The Problem of Trust

1.

Trust, I propose, should be distinguished from mere reliance, even if it essentially involves reliance.[1] Here are some examples of trust. Leaving one's closed diary on a desk where one knows one's partner will see it. Not asking for a second quote when a mechanic says one's car needs lots of work. Shaking on a deal. Following a stranger's directions. Purchasing a good that will be delivered later. And these are also examples of reliance. The difference is that although reliance can be willing – one can be confident that someone or thing will prove reliable – it can also be forced. By contrast, trust, Bernard Williams observes, "involves the willingness of one party to rely on another to act in certain ways" (2002, 88). This willingness is essential because to trust someone is to take an optimistic view of that person and their motivations. It is "to lower one's guard, *to refrain from taking precautions against an interaction partner*, even when the other, because of opportunism or incompetence, could act in a way that might seem to justify precautions" (Elster 2007, 344). Such precautions might include keeping one's diary in a locked desk, and consolidating the handshake with a legally enforceable contract. To trust is to forgo precaution and garner reassurance just from the thought that the other is trustworthy.

The attitude of trust is then part identified by this thought. We can 'trust', or so we say, our car to start or our alarm to go off. But cars and alarms aren't trustworthy and we don't, properly speaking, trust them. Rather, this is just reliance: we rely on them and do so believing that they are, or will prove, reliable. Reliability is a feature of trustworthiness but does not equate to it. Similarly, and famously, Kant's neighbours could rely on his punctuality in the same way that they could rely on a clock; they could 'trust' Kant to walk by the window at eight in the morning. But *trusting* Kant to do this, Russell Hardin argues, would require his "having their interests at heart in deciding to take his walk" (2002, 5). It would require thinking of Kant's walk as a trustworthy action, or one done, at least in part, in response to the dependence exhibited by his neighbour's 'trust'. But of course Kant did not take his walk for this reason and his neighbour's did not think he did. There might have been reliance and reliability but there was no trust or trustworthiness.

[1] See Baier (1986).

Trustworthiness is more than reliability in that it is evaluative as well as descriptive. The trustworthy act, like the opportunist one, is done in response to someone's depending on one in certain ways, but unlike the opportunist act, it is the appropriate thing to do – where 'appropriate' here amounts to a quasi-moral evaluation. Thus, we think, you shouldn't read another's private diary or give an inflated quote, and you should honour agreements and help strangers who ask for directions. The thought that the person one interacts with is trustworthy, or will prove so, then provides rational reassurance: there is no need for precaution, if this thought is true.

So trust is unproblematic if one knows that the trusted is trustworthy. But, arguably, it becomes problematic as our epistemic grounds for judging that the trusted is trustworthy lessen. However, the resulting *problem of trust* is fundamentally practical rather than epistemic: it is a problem of the rationality of reliance broadly – a problem of the rationality of cooperation. Interactions that potentially raise the issue are those that have a worst outcome: one delivers and the other defaults. One leaves one's diary on the desk and it gets read, one doesn't get a second quote and is stung, and so on. So precautions can seem justifiable. And what one seeks to do in taking precaution is to remove the problematic reliance. But if the reliance cannot be removed, what would make it rational are any grounds that provide reassurance that the other is, or would prove, reliable or cooperative? Of course, knowledge that the person is trustworthy would do this, but reliance could be rationally sustained on slighter grounds. In respect of this, Williams (1988, 118) identifies four general motivations people can have to cooperate or rely on one another. Fear of sanctions is one; the others are: particular self-interest, a positive evaluation of cooperation and a positive evaluation of friendly relations. Sanctions can be informal or formal. Informal sanctions consist in some form of social exclusion, such as loss of reputation or simple ostracism, and the further losses this causes. Formal sanctions can be anything from a fine to a prison sentence; professionals who fail in their role, for instance, can lose their legal right to practice.[2] We are sensitive to all these motivating reasons for cooperation. Moreover, we also engage with one another in ways that provide some grounds for empirical judgment of motivation. Certain features of an individual or their behavior provide what Elster (2007, 347) respectively calls 'signs' or 'signals' that allow a judgment of reliability. The *problem of trust* is then this: even if the net is cast widely, so that all that is needed to rationalize reliance is some grounds for judging reliability, we can still lack grounds for this judgement.

[2] See Blais (1987).

2.

This problem can be outlined by reference to an experiment called *The Trust Game*.[3] In this game there are two players: a trusting party or 'investor', X and a trusted party or 'trustee', Y. In the simplest version X has a certain endowment, say £100, and option of keeping this or transferring part or all of it to Y and keeping whatever remains. What gets transferred gets multiplied, say by a factor of 4, and Y then has the option of keeping the resulting sum or making a back-transfer of part or all of it and keeping the remainder. In this case, the best X could do, would be to give Y everything and hope Y splits his £400 windfall. If Y does share, both do well and the game has resulted in a cooperative outcome. But how would it be in Y's interest to return anything?

One way in which it might be is if the game were iterated. In such a situation, defection would be repaid by a lack of future cooperation, and one would thereby lose the benefits of this cooperation. So X's interest in receiving a back-transfer is also in Y's interest since it allows for another exchange. And this, Russell Hardin proposes, is all that trust amounts to: it is merely a matter of *encapsulated interest*.[4] The problem with this view is that X's interests can be encapsulated in Y's interests *only insofar as* the exchange is on-going. If the relationship is one-off, then there is no value to be got from acting as if it were on-going.[5] But then considering such cases, why think that Y will cooperate? And here we return to the motivations that people can have to be cooperative, or to rely on one another, and our knowledge of these. However, suppose now that X's position is one of ignorance; that is, suppose – as is meant to be the case in the experimental set up – that X knows nothing about Y and so is ignorant of what does or might move Y. Once this assumption of ignorance is added it becomes hard to see how anything other than defection from the start could be the rational thing to do. Admittedly, we are never entirely ignorant – even to presume that Y prefers more money to less is a presumption about preference. However, once the rationality of reliance is grounded on the judgement that the other party will prove reliable, then it seems as though reliance becomes irrational when the grounds for such a judgement are not available. In the face of ignorance, it is surely better to keep what monies one has.

This *problem of trust* is then a *sceptical problem* insofar as the facts that generate it commonly hold true; that is, insofar as (i) we need to rely on another but recognize that doing so could have a worst outcome – we rely and other proves unreliable – and (ii) we know that this interaction is one-off (or

[3] See Glaeser et al. (2000)

[4] "I trust you because I think it is in your interest to attend to my interests in the relevant matter." (Hardin 2002, 4). On the view of trust being presupposed here, this is insufficient because the judgement of trustworthiness that part constitutes it is non-evaluative.

[5] Equally, if it is not one-off but has a determinable end, once a final interaction is established the rationale for cooperation unravels by backwards induction.

one of determinate number), and (iii) we are entirely ignorant of the other's individual motivations, but recognize a general motivation to be unreliable. Given these facts, reliance would seem to be irrational, and if it is, then *afortiori* so too is trust. Yet presently, and thankfully, trust is part of the fabric of our society: thus a key experimental result is that people, with some interesting qualifications, do by and large cooperate in the Trust Game even though it is played as a one-off game under conditions of ignorance, and hold trustees to the expectation that they *ought to* cooperate in return.[6] Thus, as with other sceptical problems, the issue is how to reconcile the philosophical result that reliance seems frequently irrational with the everyday fact that trust is pervasive. The difficult question is to account for how such reliance is rational under these conditions.

3.

In practice, it is simple to render reliance unproblematic: any potentially problematic situation, one described by (i), can be altered such that facts (ii) and (iii) no longer hold of it; that is, devices can be introduced which through adding future interactions or some other means, such as sanctions, give the trustee known reason to be cooperative. But, David Gauthier observes, to solve the problem of trust in this way is simply to "miss the point": what is needed is an account of how the mere "mutual advantageousness" of cooperation, the fact that both investor and trustee benefit from it, renders it rational. Thus Gauthier distinguishes, what might be called, *internal* and *external solutions*. An external solution establishes the rationality of reliance, in a situation where (i) is true by falsifying (ii) or (iii). An internal solution shows how reliance is rational even given the truth of (i), (ii) and (iii). And he seeks to provide an internal solution.

The solution, Gauthier suggests, is to abandon the conception of rationality presupposed in generating the problem of trust. What is presupposed is a *maximising conception of rationality*: individuals pursue what is most in their interest. So where (i), (ii) and (iii) hold, and the investor is ignorant of the trustee's particular interests, he can only think that the trustee would keep all the monies if he had the chance. However, rather than see the rational outcome as an equilibrium point – the product of maximising choices – the rational outcome, Gauthier (2013, 195-6), proposes, is the *optimal* one, or that outcome that could not be better for one subject without being worse for others. The rational thing to do is to adopt a strategy that optimizes rather than maximizes, where to do this is to be what Gauthier (1986, 167) calls a 'constrained maximizer' as opposed to a 'straight-forward maximizer'. A constrained maximizer pays attention to the *interests* of others, not merely

[6] See Glaeser et al. (2000).

their *choices*, or at least does so conditional on others reasoning likewise. And it is rational, Gauthier argues, to be a constrained maximizer because one does best this way. The reason for this is simple: we have a basic disposition to cooperate only on terms that are fair and, as such, constrained maximizers have more opportunity for cooperation. "Straightforward maximizers are disposed to take advantage of their fellows should the opportunity arise; knowing this, their fellows would prevent such opportunity arising" (Gauthier 1986, 173). So it is rational for the investor to rely on the trustee to return half the windfall and for the trustee to make this fair return because were either to have the disposition to act otherwise they would have less opportunity for cooperation in general. Any gain to be made in this case would be offset by the greater loss that this exclusion would entail.

Allow that Gauthier has established the rationality of being a constrained maximizer. A constrained maximizer will cooperate and so rely on another only if the other is equally judged to be constrained maximizer. As such the straightforward maximizer "must seek to appear trustworthy, an upholder of his agreements" (Gauthier 1986, 173). And constrained maximizers must cultivate "their ability to distinguish sincere cooperators from insincere ones" (Gauthier 1986, 181). A difficulty, then, for this solution to the *problem of trust* is that it must presume this constrained maximizer ability outstrips the ability of straightforward maximizers to beguile. And this presumption just seems to be the denial of (iii): the rationality of reliance thereby hinges on a belief about the trustee's motivational dispositions, namely that the trustee is a constrained maximizer, not a straightforward one, *and recognised as such.* Thus in a situation of ignorance, were (iii) true, the rational thing to do would surely be to presume that the other is a straightforward maximizer and avoid relying on them.

Not necessarily argues Gauthier. The inability to detect straightforward maximizers would not matter if there were not so many of them. In demonstration of this Gauthier offers a sophisticated account of how the rationality of being a constrained maximizer would be preserved *if* the proportion of constrained maximizers in the population were significantly large.[7] However, the issue then becomes what reason we have for thinking that our population is like that. And the problem is that *if* our reason for thinking this is empirical, Gauthier's argument for the rationality of cooperation would seem to rest, at base, on the denial of (iii): our situation is not really one of ignorance if we know enough to empirically justify the assumption that the other is a constrained maximizer. However, arguably we do not have any *apriori* reason for making this assumption.

[7] "The more constrained maximisers there are, the greater the risks a constrained maximiser may rationally accept of failed cooperation and exploitation" (Gauthier 1986, 176).

To argue this, consider the case of Blue Jays and Viceroys.[8] Although they are good for Blue Jays to eat, Viceroys look like Monarchs that are not. This mimicry is an evolutionary lie. The Monarch signals truly that it is a poor meal, the Viceroy mimics this signal to communicate something false. The position of an investor is then analogous to the position of a Blue Jay faced with a brightly colored butterfly. Just as the Blue Jay gains from eating the butterfly if it is a Viceroy, and loses if it is a Monarch; so the investor gains from making the transfer if the trustee is a constrained maximizer and loses if trustee is a straightforward maximizer. And the ability of investors to discriminate constrained maximizers from straightforward maximizers, like the ability of Blue Jays to discriminate Viceroys from Monarchs, is limited. Suppose then a best case starting state: this limited discriminator capacity is unproblematic because there is no deception. Only Monarchs are colourful, only constrained maximizers purport to be so, Blue Jays eat all drab butterflies and the worst case scenario is never realized for the investor. From this starting point, deception will evolve: colourful Viceroys and straightforward maximizers who learn "to appear trustworthy", will flourish. But as levels of deception rise, levels of reliance go down; but they should decrease only so far, because if reliance vanished, there would be no gain to be had from deception. So at what point does the population reach a stable equilibrium? "Interestingly enough", Eliot Sober (1994, 80) observes, "it turns out that the stability of the system depends on the type of dynamics one assumes." What this means is that there is no *apriori* reason for thinking that constrained maximizers predominate. They might, just as most colorful butterflies might be Monarchs, but if this is so, this "stability depends on some highly contingent properties of the evolutionary process" (Sober 1994, 80). Which is to say it depends on some highly contingent properties of our society. So the rationality of reliance has to be ultimately grounded on an empirical belief about our social situation. This might be a solution – and in some ways it is close to the one I sketch below – but so far it is not the internal one hoped for.

4.

Philip Pettit presents a different internal solution to the problem of trust. Take any potential engagement where (i) is true, such as the Trust Game; the standard reason that an investor might rely on a trustee to share their windfall, Pettit argues, would be the judgement that the trustee is trustworthy (where by this Pettit means little more than 'is reliable'). And the standard reason one has for this judgement is the further belief that the trustee is motivated by *loyalty*, *virtue* or *prudence*, where these motivations need not be exclusive. So what grounds are then available for this judgement if the

[8] See Sober (1994).

investor is ignorant of the trustee's particular motivations? That is, if facts (ii) and (iii) hold. To answer this question, Pettit (2002, 353) suggests that one needs to recognise that people are motivated by both "action dependent goods", such things as commodities and money, and "attitude dependent goods" or such things as "being loved, being liked, being acknowledged, being respected, being admired, and so on." Moreover, the desire for these goods is *universal* in that it can be assumed "that each of us desires the good opinion of others" Pettit (2002, 354). However, if this is the case, there are grounds for thinking that the trustee will prove reliable, and so grounds for rationally relying on the trustee, even given the truth of (i), (ii) and (iii).

Pettit lays out these grounds with the following argument.

> 1. There are situations where an act of trust will signal to a trustee, and to witnesses, that the trustor believes in or presumes on the trustworthiness of the trustee ... and so thinks well of him to that extent.
>
> 2. The trustee is likely to have a desire, intrinsic or instrumental, for the good opinion of the trustor and of witnesses to the act of trust.
>
> 3. The desire for that good opinion will tend to give the trustee reason to act in the way in which the trustor relies on him to act.
>
> *Conclusion.* And so the trustor, recognizing these facts may have a reason to trust someone, even when he actually has no reason to believe in the other's pre-existing trustworthiness (Pettit 2002, 357).

As a statement of a reason we can have for relying on others, this conclusion is well observed. But as a solution to the problem of trust, the argument faces a couple of difficulties. First, to resolve the problem of trust, the desire for the good opinion of others needs to be universal in the sense that it can be taken to be a component in every interaction. Thus, premise 2 is too weak as stated with the qualifier '*likely* to have a desire' and the conclusion correspondingly so with the qualifier '*may* have a reason'. To strengthen premise 2, it would have to supposed that the desire for the good opinion of others were intrinsic rather than instrumental, since it would not be plausible to claim either that the good opinion of others was instrumentally valuable in every case, or that it was instrumentally valuable in every case where the trustor didn't have an intrinsic care for the good opinion of the trustee. Suppose then that this is true: in every case the trustee will have an intrinsic desire for the good opinion of the trustor and of witnesses to the act of trust. Even with premise 2 thus strengthened, it is still insufficient for the trustor to have a reason to trust – that is, rely – in the moot cases. The issue now is that even if the trustee always has a desire for the good opinion of others, and so a reason to be 'trustworthy', this reason need not suffice for being so. This is because in most cases the trustee will have more interests than this at play in the

interaction. So what needs also to be true is that the trustor has grounds for thinking that this desire for good opinion of others is either the only operative desire or the dominant one. And while the trustor could well have grounds for thinking this, the trustor would not have such grounds when the interaction takes places under conditions of ignorance; that is, when (iii) is true. Put another way, if (iii) is true, the trustor lacks the unique reason for relying on the trustee that Pettit hypothesizes. So the hypothesis of this reason offers no internal solution to the *problem of trust*. Again it depends on particular empirical belief whose possession implies that one of the facts generating the problem does not hold.

5.

In the remains of this paper, I hope to elaborate on a solution to the *problem of trust* suggested by Bernard Williams. In this section, I outline this solution.

As observed, Williams (2002, 88) takes trust "in its most basic sense" to be no more than "the willingness of one party to rely on another to act in certain ways" This willingness implies some expectation of motive but "it does not imply that those motives have to be of some specific kind" (Williams 2002, 88). In particular, it does not imply that the motives of the trusted are those of the trustworthy person: the judgement that the trusted was sensitive to the possibility of punitive sanctions, for instance, might suffice for the prediction that the trusted would prove reliable, and so rationalize reliance. However, Williams goes on to argue, such judgements of motive do not resolve what I've called the problem of trust. Williams's interest in this problem lies with a particular instantiation of it: what motives we have for telling others the truth, or being sincere. Any functioning society will involve cooperative engagements which demand information be communicated between individuals. So sincerity is clearly desirable from the social point of view, but it is equally not always in an individual's best interest. For instance Williams imagines the case of the hunter who has just made a kill, which he would prefer to keep for himself and his family. Sincerity is not in this hunter's interest. The problem this case illustrates is that:

> The value that attaches to any given person's having this disposition [Sincerity] seems, so far as we have gone, largely a value for other people. It may obviously be useful for an individual to have the benefits of other people's correct information, and not useful to him that they should benefit of his. So this is a classic example of the "free-rider" situation (Williams 2002, 58).

Whilst it is always in an audience's interest to be informed, sincerity need not best serve a speaker's interest and as audiences we know that this is the case. So, what I have called, the problem of trust threatens. The source of this

problem, Williams argues, is that sincerity – trustworthiness in speech – has only been given *instrumental value*: its value is given by that good which follows from information being pooled, such as the good brought about by cooperative endeavor. And when it is only valued in this way there will always be the possibility of a fissure between interacting parties' interests such that trust exposes the trusting, or those who have a trustworthy disposition, to 'free-riders'. An adequate solution must not allow the realization of this possibility to be systemic. What this shows, Williams (2002, 59) claims, is "that no society can get by … with a purely instrumental conception of the values of truth" of which sincerity, which is trustworthiness in speech, is one. What any society thereby requires is that such trustworthines be *intrinsically valued*, so that it is thought to be "a good thing (many other things being equal) to act as a trustworthy person acts, just because that is the kind of action it is" (Williams 2002, 90).

Three points need to be made about this idea of intrinsic value. First, to value trusting and trustworthy acts intrinsically is to take an evaluative stance: it is *a good thing* to act as the trusting and trustworthy person acts, which is to say that one *ought* to act this way. So if solving the problem of trust requires thinking of trust in this way, trust must be taken to involve more than mere reliance – as claimed in section 1. Second, it follows, and Williams stresses this implication, that there can be no solution to the problem of trust that starts, as Gauthier's and Pettit's do, from an individual's interests. This is because no argument from such a starting point, namely one "which sets out from a game-theoretical formulation of the problems of trust could possibly show that trustworthiness had an intrinsic value" (Williams 2002, 90). The most that could be concluded is that it can be in one's interest to act as if trust and trustworthiness had intrinsic value; but this is consistent with pretending they had this value, rather than valuing them, and so would not eliminate the issue of free-riders or solve the problem. However, third, talk of intrinsic value raises a dilemma. On the one hand, intrinsic value seems rather mysterious, amounting to the claim that "there is nothing else to be said about its [trust's] valuableness – it is good because it is good, and that is all there is to be said about it" (Williams 2002, 90). On the other, if some account of its value is offered "such as securing cooperative activity which is in everyone's interest" (Williams 2002, 90) then it seems as though one is giving a reductive account: its value is instrumental after all.

What is needed is a non-reductive explanation of what it is for something to have intrinsic value. And a sufficient condition for this, Williams proposes, is that

> first, it is necessary (or nearly necessary) for basic human purposes and needs that human beings should treat it as an intrinsic good; and, second they can coherently treat it as an intrinsic good (Williams 2002, 92).

A genealogy, he then claims, satisfies both of these conditions. Sincerity is 'necessary (or nearly necessary) for basic human purposes' in that were individuals not to have the disposition to be sincere, or trustworthy in speech, no cooperative endeavour would be possible and there would be no such thing as society, at least as we know it. Except this claim is too strong as stated: cooperation in communication could be established on the same grounds as cooperation more generally: through knowledge of the motivations we know people can have to cooperate. But what could be claimed is that it is only through individuals having the disposition to be trustworthy that cooperation could be sustainable *once conditions (i), (ii) and (iii) are in place*. So trustworthiness's having intrinsic value is necessary for what I called an *internal* solution to the *problem of trust*. And a genealogy then shows how sincerity 'can coherently be treated as an intrinsic good' through showing how our having the disposition to be trustworthy "makes sense to [us] from the inside, so to speak" (Williams 2002, 92). And this is what I hope to do in the remains of this paper: to offer a conceptual analysis of trust and trustworthiness – an account of our understanding of these notions – that makes sense of our being motivated to act in trusting and trustworthy ways just because we value acting in these ways, where this motivation can be undimmed by an interaction occurring under conditions (i), (ii) and (iii).

However, before continuing with this analysis, it is worth observing that the kind of generalised disposition to trust and be trustworthy, which follows our intrinsically valuing trust and trustworthiness, is quite different to the dispositions to 'trust' and be 'trustworthy' as characterised by Gauthier and Pettit. The rational disposition to have, or strategy to follow, according to Gauthier is to be a constrained maximizer when and only when one judges one's interaction partner is so and to be a straightforward maximizer otherwise. To be a straightforward maximizer would be to be untrustworthy; it would demonstrate a failure to give the appropriate deliberative weight to a trusting party's interests. Thus, Gauthier's proposal suggests that one can be motivated to be trustworthy in some situations and motivated to be untrustworthy in others, which implies that the disposition to be trustworthy can be switched on and off in response to the situation one finds oneself in.[9] However, it is not plausible to suppose that a generalized disposition to trust or be trustworthy can be conditional in this way. Take the Trust Game. Were an investor to have a generalized disposition to trust, the investor would be disposed to make a transfer irrespective of any judgement of the trustee's likely response. To simply keep the money would seem the 'wrong' thing to do. Of course, this is not say that there is no knowledge that would stop the investor from making a transfer, but it is to suggest that the disposition to do so cannot be conditional in the way Gauthier describes. Rather, in having this

[9] The same thing is implied by a strategy of tit-for-tat which would require one's trustworthiness be conditional on the previous play of one's interaction partner.

disposition the investor would see himself as having a reason to make a transfer, a reason that in some circumstances might be unfortunately outweighed, but a reason irrespective of these circumstances not a reason that only exists conditional upon them.

Sticking with the Trust Game, the rationalisation of trust proposed by Pettit (in the strengthened form suggested) proceeds from the premise that we have a basic desire for the good opinion of others. While this may be true, both of the trustee and in general, the trustworthy response to the investor's transfer does not start deliberation from the fact that a return transfer will receive plaudits. Rather, the trustworthy person starts deliberation from the fact that the investor trusted or made a transfer, and is consequently in a position of vulnerability. And were the investor to have a generalised disposition to trust, this disposition would, in part, be that of presuming the trustee would start deliberation from this fact. To think otherwise, and, in particular, to suppose that the trustee deliberates from the premise that a return transfer would garner praise would be to think badly of the trustee; it would be to think that their motivations are not 'right', or not those of the trustworthy person.[10]

6.

The challenge, as Williams puts it, is to give an "explanation without reduction"; to say why trust and trustworthiness are valued without making this value instrumental. This challenge, I now want to suggest, cannot be met all the while trust is thought of as a three-place predicate whose instances are X trusting Y to φ. While trust can be three-place, we trust one another to act in various ways – not to read our diary, to give a fair quote and so on – this form of trust, its *contractual* form one might call it, cannot be fundamental; or at least it cannot be so, if trust is something that is intrinsically valued. For the value, for X, in trusting Y to φ is fundamentally the (action dependent) goods that come from Y φ-ing. Its value is fundamentally instrumental (the value attached to keeping one's indiscretions hidden, or paying a fair price, for instance). Any value intrinsic to trust must adhere simply to X's having the trusting attitude; that is, in X's trusting Y, or just in X's trusting. That is to say, insofar as trust is intrinsically valued, it fundamental form must be, as one might say, *attitudinal*; it must be two-place, or even one-place but not three. It is then X's having a generalised disposition to trust in this way – to simply trust, or to trust Y – that explains why it is that X trusts Y to φ.

Focussing, then, on attitudinal trust, what is this generalised disposition to trust? It is, I propose, the disposition to presume that one's

[10] Thus, there is something oddly self-defeating about the desire for good-opinion. Compare (Elster 2007, 351).

interaction partner, the trustee in the Trust Game for instance, is trustworthy. This then raises the question of what it is to be trustworthy. Here Williams observes

> one thing it needs to be is the disposition of an agent to be reliable, not in the sense that you can rely on him to help you (that is a different disposition, helpfulness), but in the sense that he will help you if he has told you that he will help you or, perhaps, if he has led you to believe that he will (Williams 2002, 92).

The answer I think is much closer to the one Williams rejects. His focus is wrongly on trust in its three-place or contractual form, and he understands the trustworthy disposition by reference to this form. But if it is the attitudinal form that the disposition of trustworthiness needs to be understood by reference to, then trustworthiness is much closer to helpfulness in that it is essentially a benevolent attitude, or an attitude of goodwill, that is manifest by an appropriate response to another's dependence. Appropriate in the sense of right: the trustworthy person can be relied on to do 'the right thing'. It is this benevolent attitude and associated goodwill that then allows for contract.

That these attitudinal forms of trust and trustworthiness are fundamental can then be supported, I think, by five independent bits of evidence.[11]

The first piece of evidence comes from everyday language. First, both the two-place predicate 'X trusts Y' and the three-place 'X trusts Y to φ' have unique and irreducible meanings. It is true that sometimes we use 'X trusts Y' as shorthand for 'X trusts Y in some particular way'; for instance, asked why she left her diary visible on the desk, X might reply that she trusts Y, and by this mean that she trusts Y not to read it. However, this is not the most straightforward use of 'X trusts Y', which is that of a description of X's attitude towards Y as a trusting or trustful one.[12] And by implication that Y, the object of X's attitude of trust, is someone who can be trusted. By contrast, 'X trusts Y to φ' is a metaphysically hybrid notion (like knowledge, as opposed to belief, is ordinarily understood) in that it describes an action – X's relying on Y to φ – and says of that action, that it is done with a certain attitude, which is best described as trustful. That is, it reports the fact of X's reliance and X's attitude to relying; it is not a direct description of X's attitude and does not carry the implication that Y is someone to be trusted. Now while 'X trusts Y' might imply a disposition to rely on Y in various ways, and one to rely on Y to φ, it cannot be reduced to such a disposition and formalised as '∀φ, X trusts Y to φ'. For 'X trusts Y' might be true, while '∀φ, X trusts Y to φ'

[11] For further argument see (Domenicucci and Holton 2016).
[12] See (Becker 1996, 44-5).

will almost certainly be false: there is always a limit to what we will trust others to do. Moreover, this does not seem to be merely a quantification issue, since there is no restricted range of φ, R, for which '∀φ∈R, X trusts Y to φ' stands as an adequate formalism of 'X trusts Y'. For while it might be true that a complete lack of willingness to rely on Y would falsify the claim that 'X trusts Y', there is no particular way in which X must rely on Y for this claim to be true. However, while 'X trusts Y' and 'X trusts Y to φ' are unique statements, there is some implication from the former to the latter but not vica versa. If X does trust Y, then there must be some φ for which X trusts Y to φ. But that X trusts Y to φ does not, in any way, imply that X trusts Y more generally, even if this would often also be true. Thus, of the two predicates, the two-place one is arguably more fundamental.

Similar things may then be said when comparing the one-place predicate 'X trusts' with the two-place predicate 'X trusts Y'. The former equally seems to have a place in everyday language: "we *do*", Uslaner (2002, 22) observes, "speak of 'trusting people' generally". And this form does not seem reducible to '∀Y, X trusts Y' for similar reasons. It will not be that X trusts *everyone*, and there is no determinate range of people that X's trust must range over. Rather, 'X trusts' seems to make a different claim: that X has faith in people, in some "generalised other", as Uslaner (2002, 24) says, not faith in any specific person or description. But again, while 'X trusts' and 'X trusts Y' seem to be different and unique statements, there is some implication from the former to the latter but not vica versa. If X trusts, there must be some Y that X trusts, but that X trusts Y does not in any way imply that X trusts more generally. So of the two predicates, the one-place one is arguably more fundamental. Thus, the heart of our notion of trust seems to be simply an attitude of trust, which may, but need not, take specific persons as its object, and which can support, but need not, the act of relying on persons.

The second and third bits of evidence come from considering trust in conjunction with distrust. "To understand trust", Katherine Hawley (2014, 1) says, "we must also understand distrust, yet distrust is usually treated as a mere afterthought, or mistakenly equated with an absence of trust." The mere absence of trust might report nothing about one's attitudes but rather stem from the fact that there is no cause for reliance. The car mechanic I trust when I don't seek a second quote, I don't trust to deliver my mail. However, my lack of trust here is its mere absence: I don't trust the mechanic in this regard not because I don't think him up to the job but because that is not his job. So I don't rely on him in this respect. Distrust, however, is not the mere absence of trust: it is an attitude in its own right, and one might expect as Hawley (2014, 4) proposes, there to be analytic connections between the attitudes of trust and distrust; such as, for instance, that if distrust is an appropriate attitude to take, then trust is not. However, that there are such analytic connections is hard to maintain if the fundamental notion of trust is

taken to be three-place or 'X trusting Y to φ'. Given that trust in this sense is metaphysically hybrid, any failure of trust can always be down to the failure of the action component. (My not trusting my mechanic to deliver my mail because I don't rely on him to do this.) But then trust could be inappropriate because of some inappropriateness in this action component; it would, for instance, be wrong to trust my mechanic to deliver my mail. However, this wrongness does not imply it is right to distrust my mechanic. So to keep the parallel between trust and distrust, the focus needs to be on the attitudinal conception of trust: trusting Y and distrusting Y. Moreover, this is implied by the fact that there is no three-place distrust predicate: we do not say 'X distrusts Y to φ' – I don't distrust my mechanic to deliver my mail! The absence of this predicate form then suggests that it is the two place 'X trusts Y' that is fundamental.

Trust and distrust, it is often said, are contraries but not contradictories. [13] And this is true all the while trust is conceived contractually, or as three-place. In this case, a lack of trust need not imply distrust because there might be a lack of trust because there is a lack of reliance; there is no contract, as it were, or *commitment* as Hawley (2014, 10) would say. However, a lack of trust can imply distrust. Where trust is the background attitude – where it is two-place or one-place – if trust is lost what remains is not merely its lack but distrust. Suppose X trusts Y. This trust is manifest in X's disposition to rely on Y in various ways. And were X to report that he trusts Y, what X would thereby describe is an attitude towards Y that is a basic attitude that one can take towards a person, which involves making positive presumptions about their goodwill towards oneself. Remove these positive presumptions, so that it can no longer be taken for granted that Y will act in certain ways and will not act in others and what is left is distrust. For example, you might not seek a second quote simply because you trust your mechanic, and if so, you just presume the quote is honest; you might leave your diary lying on the desk simply because you trust your partner, and if so, so you just presume they won't read it; or suppose you trust your partner, if so you will just presume they are not cheating on you; and so on. Remove trust in these cases, so you no longer presume the quote honest, the diary safe or your partner faithful and these situations are now ones of distrust.

Relatedly, we tend not to trust people *not to do things*. For instance, you don't trust your partner *not to have an affair*, not because they can't be trusted in this but because such trust is peculiarly self-defeating. To trust them not to have an affair would be to draw their attention to the fact that you do not presume they will not, which amounts to not trusting them in this respect. Equally, your partner would not reassure you were they to say 'don't worry I won't be unfaithful'. This should be unspoken, part of what is

[13] See (Jones 1996, 15).

presumed by mutual trust. The same goes for one-place trust, in having a non-directed attitude of trust we presume things about how people in general will behave towards us. For instance, we presume they won't be unpromptedly aggressive. This presumption, Williams observes, can be sustained by reasoning "in desperate circumstances", but in "better times" we just take it for granted. And it needs to be taken for granted because "[o]ne is not likely to be reassured by someone who says, 'I promise not to murder you'" (Williams 2002, 89). Thus a proper account of the relation of trust to distrust, and the recognition that these can be contradictory, requires a purely attitudinal conception of trust. And it is then hard to see how it is not this attitude that is, as Williams (2002, 88) says, the basic form of trust "on which all social interaction depends".

The fourth piece of evidence concerns the relationship between trust and trustworthiness. The attitude of trust is, in part, identified by its relation to the thought that the trusted is trustworthy. Reassurance comes from this thought. However, this connection between trust and trustworthiness is broken if trust is conceived contractually, or as three-place. Under this conception, say in a case where X trusts Y to φ, the thought that Y is trustworthy is, at least, that Y will reliably φ. However, it might be that this is not the trustworthy thing to do and, indeed, can be quite the opposite. This might be illustrated by a case where the trusting party is in error. Imagine a hot parched land and X arriving thirsty at Y's homestead. He asks Y for water from the well that stands in front on Y's house, and Y responds by telling him that he can't have that water and then goes inside. In fact Y has gone to fetch X some clean water from the tank at the back of the house, the water in the well standing at front of the house having been poisoned by livestock that fell into it and died at the start of Spring. Not knowing this, X will judge Y untrustworthy, *and if trustworthiness is identified by reference to trust* Y would be so. But of course, Y's response is the right and trustworthy one. This point is made and developed by Knud Ejler Løgstrup in his discussion of trust.

> The other person's interpretation of the implication of the trust offered [that is, the trusting party Y's interpretation] ... is one thing, and the demand which is implicit in that trust ... which I must interpret is quite another thing (Løgstrup 1997, 21).

Responding to trust cannot be "merely a matter of fulfilling the other person's expectations and granting his or her wishes" (Løgstrup 1997, 21). This is because, in the trust situation, such as that of the poisoned well, "what we are speaking of is a demand for love, not for indulgence" (Løgstrup 1997, 21). Thus the demand on the trusted – what Løgstrup calls the *radical ethical demand* and might be called the demand that X *be trustworthy* – is generated by the fact of the trusting party's dependence.[14] It is not generated by X's *attitudes* –

[14] Although Løgstrup doesn't ever talk of trustworthiness, see (Faulkner 2016).

that is, by his trust. But this is to say that trustworthiness cannot be defined with respect to trust if trust is conceived contractually, or as three-place. The analytical connection between trust and trustworthiness is preserved if trust is taken to be merely an attitude, or as two-place. For suppose, in the poisoned well case, that X simply trusts Y. In trusting Y, X will think that Y is trustworthy. And in thinking this, X will not place any specific expectation on Y, but will rather just expect it of Y that Y does the right or trustworthy thing.

Connected to this point is Katherine Hawley's observation that trust can be unwanted. In this regard she gives the example of trusting her colleagues to buy her champagne, in a situation where, for whatever reason, she is to be honoured. Now it might be that her colleagues plan to buy her champagne but, Hawley (2014, 7) observes, "[s]till, they do not invite or welcome my trust in this respect; instead, they want to give me a treat, not merely to act as trustworthiness requires, and certainly not to risk betraying me if they forget to buy the champagne". This observation is good, but her trust is unwanted, in part, I suggest, because it implies the *falsehood* that the colleagues would be untrustworthy if they did not supply it. This is false precisely because being trustworthy is not a matter of wish fulfilment. It is a matter of doing the appropriate thing, which might still be to buy champagne, and Hawley's colleagues want to be trusted to do this. So to trust them is to trust in the two-place sense, and such trust would not be unwanted. What is objectionable is the implicit contract, not the background attitude.

The fifth and final piece of evidence for the priority of two-place over three-place trust comes from a consideration of infant trust. Any account of trust, Annette Baier (1986, 244) proposes, should accommodate infant trust. And this generates the constraint "that it not make essential to trusting the use of concepts or abilities which a child cannot be reasonably believed to possess". Suppose now that X trusts Y to φ. In trusting Y to φ, X will take an optimistic view of Y and her motivations; and in so taking this view X will, at the very least, presume that Y will φ, and φ because X manifestly depends on her doing so. To do what X trusts for this reason would be to be trustworthy. So in trusting Y to φ, X presumes that Y is trustworthy (or *at the very least X presumes this*: often, if not ordinarily, this presumption will be an item of knowledge or firm belief). Thus X's trusting Y to φ involves a complex of reasoning. It involves imagining the trust situation from Y's perspective, imagining Y's recognition of X's dependence, and imagining Y seeing this as a reason to do what X depends on Y doing. Now it is arguable that this kind of second personal reasoning is both prosaic and fundamental to moral thought.[15] However, it is not the kind of reasoning that an infant could engage in. By contrast, suppose that X trusts Y; for instance, an infant X trusts his mother Y. In trusting his mother, X will have the thought that she is

[15] See (Darwall 2006) and (Faulkner 2014).

trustworthy and, at the very least, presume this thought to be true. But this thought need not be articulated in these terms, and its possession involves no second-personal awareness; it is merely the thought that his mother will do the right thing, had in a context of dependence. And even this thought need not be articulated: it amounts to no more than confidence or faith in his mother's actions. This does seem to be the kind of thought that an infant could have. Suppose then that Baier's constraint on accounts of trust is plausible. What this implies is either than an account of trust must satisfy her constraint, or acknowledge that two-place trust is a more basic form than three-place.

7.

Where does this leave the problem of trust? This problem, I argued in section two, is confronted when three facts hold true of a situation where one might rely on someone: (i) reliance has a worst case outcome – one relies and other proves unreliable; (ii) both parties know the interaction to be one-off, or of a determinate number; and (iii) one is entirely ignorant of the other party's motivations, but know there is some general reason to be unreliable. An *internal* solution to this problem is then one that explains how it is that reliance is rational in such situations with a worst case outcome even when (ii) and (iii) hold. An internal solution is what is philosophically wanted because we do rely in these circumstances, and any other solution would fail to make sense of the rationality of doing so. Both Gauthier and Pettit advance internal solutions, which for the reasons discussed, I think fail. And this paper advances an internal solution, which might be called the *trust-based solution.* According to this solution, the problem of trust is insinuated by the background assumption that it is the subject's beliefs and desires – their preferences – that explain action; and that rational action is a matter of maximizing preference satisfaction. This insinuates the problem because it excludes the reason that can, and often does, motivate reliance, which is simply the attitude of trust. Thus, with respect to the Trust Game, and the question of why rely on the trustee when (ii) and (iii) hold, the answer can be just that the investor is trusting. And this answer can be not merely a description of the investor's psychological disposition; it can also capture the investor's reasons.

This raises the question of how it is that a trusting attitude can rationalize reliance. And this question is, in fact, two: how does a trusting attitude rationalize reliance? And, why does a trusting attitude rationalize reliance? With respect to how, the answer is via what is involved in thinking well of others, which is what having a trusting attitude amounts to. Suppose that X is trusting (one-place) or specifically trusts Y (two-place). If this is true, then in a case where X has a choice of relying on Y, X will presume that Y will

do the right or trustworthy thing. In the Trust Game, this presumption would be that trustee will return a fair share of the money gained. Making this presumption is constitutive of having a trusting attitude: it taking this attitude one thinks well of others, and this presumption articulates this thought in a context of potential reliance. Thus the presumption is not based on any evidence of Y's reliability, and, as such, it would be available even if X knew nothing about Y's particular motivations. It would be available that is, in a situation where (i), (ii) and (iii) hold true. However, the question is then whether this presumption is available to the investor in the Trust Game when the investor knows that there is some general reason for the trustee to be unreliable. The answer, I think, is that it is a feature of trust that it has a certain resilience to doubt. Were X to possess knowledge of Y's particular motivations to the effect that Y's preference was always for profit, trust would become hard (though I conjecture, it might still be possible to give Y the benefit of the doubt). However, in the absence of any such particular grounds for worry, the nature of trust as an optimistic attitude, involves giving no credit to a mere general worry.[16] It follows that the presumption that Y will do the right or trustworthy thing would be available, and that, as a consequence, it would be rational for X to make the transfer. If X's attitude is trusting – if X trusts (one-place) or trusts Y (two place) – then it is rational for X to rely on Y in certain respects. In particular, if X's attitude is trusting, it is rational for X to rely on Y to make a back transfer, even where (ii) and (iii) hold true.

Why, then, does a trusting attitude rationalize reliance? The short answer: because trusting and being trustworthy are actions that we intrinsically value; and in valuing behaviours that are so described we think that, other things being equal, we have a reason to act in these ways, and acting in these ways is the right thing to do. Moreover, valuing these behaviours cannot be translated into a statement about our preferences, even if it might imply such; value is not, as Gauthier (1986, 49) says, merely "a measure of individual preference". The solution to the problem of trust is *not* this: change the preference ordering by adding a desire to be trusting and trustworthy. The preference ordering is rather kept as described: the problematic cases of reliance are ones where (i) holds true; they are cases with a worst outcome, which is that one relies and the other proves unreliable. These are cases were some reassurance would seem to be needed. Thus the solution is not to change these cases by adding such a desire to be trusting that what was the worst outcome ceases to be so. Rather, in intrinsically valuing trusting and trustworthy behaviours, we recognise reasons for acting in these ways that are not reducible to our preferences for so acting. Thus, in intrinsically valuing these behaviours, we think that rational deliberation in a potential reliance situation *should start* from the presumption that the other

[16] In this respect, one might say that trust is the default, see (Stern 2016).

party will behave trustworthily.[17] And, when trusted, we think that rational deliberation *should start* from the fact of the trusting party's need. We then explain and justify our own and others' behaviour by reference to these prescriptions. Thus the claim about intrinsic value can be understood as a claim, as Williams says, about how these actions and their accompanying attitudes make sense to us from the inside.

Moreover, deliberating as the trusting and trustworthy person cannot be captured in terms of a preference for following a norm – the norm, for instance, of trusting, or presuming that the other party is trustworthy. This is because deliberating in these ways involves *seeing the potential reliance situation in a certain light*, where this perceptual metaphor highlights the immediacy of the normative judgement. Take the Trust Game. The right, or trustworthy, way of viewing an investor's transfer is as being a reason to make a back-transfer, and this is how a trustworthy trustee would see it. Then in presuming that the trustee is trustworthy, the trusting investor would likewise see the situation generated by their transfer as one wherein the trustee has this reason. That is, the perception of this situation *will contain within it* a judgment about what ought to be done and, correlatively, about how action is to be explained.[18] There is no inference from the thought that some norm applies in this case conjoined with the desire to follow this norm; and to suppose that there is would be to miss the immediacy of what is essentially a non-inferential judgement.

In conclusion, let me make three observations about this trust-based solution. First, the claim is *not* that a preference-based explanation of action generally fails; on the contrary, this form of explanation is good in very many cases. Moreover, it can be the best way of predicting behaviour in a potential reliance situation. The trustee in the Trust Game might be Gauthier's straightforward maximiser. Now trust tends to exclude doubt.[19] So if one's attitude is trusting, then, other things being equal, one will put aside the worry that another is not similarly motivated – if this worry occurs at all. However, other things might not be equal: one might have concrete grounds for doubt. And if this were so, the only way to render reliance rational would be to support it with a belief about outcome. At this point, a calculation of preference would be needed, since an attitude of trust could not rationalize reliance give sufficient evidence that it is not mutual. However, what makes this normative claim true, ordinarily, is that given such evidence, the attitude

[17] In this respect, as Løgstrup observes, "we do not normally advance arguments and justifications for trust as we do for distrust" (Løgstrup 1997, 18). Again see (Stern 2016).
[18] See (McDowell 1978, 100-101).
[19] This point is well-made by Möllering (2009, 140), "Vulnerability is a precondition for trust ... [but] trust captures the highly optimistic expectation that vulnerability is not a problem"

of trust would no longer be available.[20] It would be replaced by an attitude of distrust. What our intrinsically valuing trust and trustworthiness then explains is why distrust is not so pervasive.

Second, the philosophical justification of our valuing trust and trustworthiness, which is Williams's genealogical argument, is not what Korsgaard (1986, 22) calls an 'ultimate justification'; that is, it does not purport to show that "all rational persons could be brought to see that they have a reason to act in the way required", which, in this case, is to act as the trusting or trustworthy person does. If Gauthier's or Pettit's justification of reliance – 'trusting' as they say – were successful, then such an ultimate justification would be available. But a significant burden of this paper has been to support Williams's contention that the problem of trust is endemic to this conception of reason. What the solution argued for here involves is an expansion of the scope of what counts as a reason for acting. Korsgaard, of course, argues for such an expansion in the hope of ultimately justifying moral principles. It would take more time to argue this point but the problem this strategy confronts, at least in application to the present case, is that it must regard strategic reasoning as a failure of rationality. But there seems no *rational deficiency* in reasoning in a tit-for-tat manner.

Third, the 'morality' of trust is thus a veneer in the sense that we need not have achieved this way of looking at things, and this way of looking at things could easily break down. In Williams's terms, we are lucky to live in "better times" Williams (2002), p.89.[21] However, this is not to make intrinsic value purely a matter of our valuing: our having this set of values is further valuable in the sense that it would be a bad thing if we saw the world otherwise. These are *better* times. In this respect, the shift in how the trust situation is conceived that is involved in having a trusting attitude is not merely a matter of having a certain upbringing, it is a matter of having a good upbringing. It is just that this notion of goodness cannot be worked into correctness since there could be other ways of securing the motivations needed to render trust non-problematic.[22][23]

[20] 'Ordinarily' because there may be cases – a parent's relation to their child, for instance – where one is bound to trust almost come what may.
[21] Equally, it is a contingent matter if truth-telling is predominant; see (Sober 1994).
[22] See (Velleman 2010) for relevant ethnographic data.
[23] This paper has had a particularly long production time and gone through many incarnations. Thanks are owed to audiences in Arizona, Bristol, Edinburgh, Griefswald, Manchester, Sheffield, Stirling, Warwick and York. Particular thanks are owed to Don Fallis, Katherine Hawley, Richard Holton, Arnon Keren, Guy Longworth, Neil Manson, Matt Nudds, and Bob Stern.

Baier, A. 1986. "Trust and Antitrust". *Ethics* 96:231-60.
Becker, Lawrence. 1996. "Trust as Noncognitive Security about Motives". *Ethics* 107:43-61.
Blais, M. 1987. "Epistemic Tit for Tat". *Journal of Philosophy* 82 (7):335-349.
Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect and Accountability*. Cambridge MA: Harvard University Press.
Domenicucci, Jacopo, and Richard Holton. 2016. "Trust as a Two-Place Relation". In *The Philosophy of Trust*, edited by P. Faulkner and T. Simpson. Oxford: Oxford University Press.
Elster, Jon. 2007. *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*. Cambridge University Press.
Faulkner, Paul. 2014. "A Virtue Theory of Testimony". *Proceedings of the Aristotelian Society* 114 (2):189-212.
———. 2016. "Trust and the Radical Ethical Demand". In *What is Ethically Demanded? Essays on Knud Ejler Løgstrup: The Ethical Demand*, edited by H. Fink and R. Stern. Notre Dame: University of Notre Dame Press.
Gauthier, David. 1986. *Morals By Agreement*. Oxford: Clarendon Press.
———. 2013. "Achieving Pareto-Optimality: Invisible Hands, Social Contracts, and Rational Deliberation". *Rationality, Markets and Morals* 4:191-204.
Glaeser, E., D. Laibson, J. Scheinkman, and C. Soutter. 2000. "Measuring Trust". *Quarterly Journal of Economics* 113:811-845.
Hardin, R. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.
Hawley, Katherine. 2014. "Trust, Distrust and Commitment". *Nous* 48 (1):1-20.
Jones, K. 1996. "Trust as an Affective Attitude". *Ethics* 107 (1):4-25.
Korsgaard, Christine. 1986. "Skepticism about Practical Reason". *The Journal of Philosophy* 83 (1):5-25.
Løgstrup, Knud Eljer. 1997. *The Ethical Demand*. Notre Dame: University of Notre Dame Press.
McDowell, John. 1978. "Are Moral Requirements Hypothetical Imperatives?". In *Mind, Value and Reality*, edited by J. McDowell. Cambridge, MA: Harvard University Press. Original edition, 1978.
Möllering, Guido. 2009. "Leaps and Lapses of Faith: Exploring the Relationship Between Trust and Deception". In *Deception: From Ancient Empires to Internet Dating*, edited by B. Harrington. Stanford: Stanford University Press.
Pettit, Philip. 2002. "The Cunning of Trust". In *Rules, Reasons, and Norms*, edited by P. Pettit. Oxford: Clarendon Press.
Sober, Elliott. 1994. "The primacy of truth-telling and the evolution of lying". In *From a Biological Point of View: Essays in Evolutionary Philosophy*, edited by E. Sober. Cambridge: CUP.
Stern, Robert. 2016. "Løgstrup on the Priority of Trust". In *The Philosophy of Trust*, edited by P. Faulkner and T. Simpson. Oxford: Oxford University Press.

Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. Cambridge: Cambridge University Press.
Velleman, David. 2010. Regarding Doing, Being Ordinary.
Williams, B. 1988. "Formal Structures and Social Reality". In *Trust: Making and Breaking Cooperative Relations*, edited by D. Gambetta. Oxford: Blackwell.
———. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.