

A Note on the Puzzle of Trust

Paul Sheehy

Introduction

We are familiar with the notion of trust. Trusting another person is part of what it is to stand in relations of love, friendship and co-operation. Trust is central to the faith characteristic of the membership of a religion. Trust is presupposed in the anguish of betrayal. It is natural to think that either someone is acquainted with trust or else she is in some profound sense deficient or lacking in a form of knowledge vital for a full sense of humanity. That deficiency, moreover, is not a source of criticism but of sympathy. For, an ignorance of or inability ever to trust is to be forced to endure a stunted, curtailed kind of engagement with a world of other persons. Trust is both commonplace and vital. In its ubiquity we perhaps lose sight of its importance until we find ourselves unable to act because we cannot trust the other(s) now or because our trust leads us to harm.

There is, then, one way at least in which there is absolutely nothing puzzling about trust. People do for the most part stand in trusting relations. Taking myself to be typical in this respect I trust certain others and I am in turn trusted by others. The puzzle arises when we turn to the question of why and how trust is possible. When a person trusts another or others she is displaying a confidence in them. Intuitively, we call this attitude *trust* when the confidence outstrips or outreaches the grounds which one might reasonably or ordinarily regard as giving rise to that confidence. On an influential view of human nature it is difficult to understand why such an attitude to others would arise as a widespread phenomenon. Strip human nature to its essential elements and we see that we are rational, maximising agents ultimately driven by self-interest.¹ To trust another is to expose oneself to the risk of betrayal; to be trusted by another and to act in a trustworthy fashion may be to forego an opportunity to seize something to one's own immediate advantage.

The puzzle is then, first, why we trust and why it is widespread if we are driven by our own self-interested concerns. A second aspect of the puzzle is whether our intuitions about the nature of trust ought to be taken at face value. Perhaps with respect to this second part you may not have any obvious intuitions; or now you turn to reflect upon them they appear confused or uncertain; or perhaps you are firm and confident in your intuitions – but why? That is the call to conceptual analysis; to get clearer about what we mean by a term and when it is apt to deploy it.² If we are to understand love, friendship and the bases of social co-operation we had better attempt to elucidate the nature and role of trust in the social world.

In the present paper I cannot hope to undertake that ambitious task. Instead I shall attempt to adumbrate one part of any response to the problem of determining why trust is possible. I consider what we *mean* by the term trust.

Defining Trust

A typical dictionary definition identifies trust as 'reliance on and confidence in the truth, worth, reliability etc. of a person or thing'.³ On this view trust is seen as a kind

of reliance. For one person to trust another is for her to believe that the other can be relied upon. That is, if I am to trust some other person, then I must have some way of assessing the risk of their failing to act in the appropriate fashion. The fundamental problem is one of overcoming ignorance about how others will act. In some circumstances there may be no real difficulty. When the robber has a gun pressed against the victim's head there is little scope for doubt about how the victim will respond to the request to hand over her money. The robber can rely on the victim. For trust to be possible amongst people, then, there has to be some basis on which they can regard each other as being sufficiently reliable. The difficulties that arise when we cannot be confident in the reliable performance of others are modelled in Hobbes' famous account of the state of nature. In the absence of an effective sovereign authority our nature brings about a situation of instability. As Hobbes observes:

So that in the nature of man, we find three principal causes of quarrel. First, competition; secondly, diffidence; thirdly, glory. The first maketh men invade for gain; the second, for safety; and the third, for reputation. The first use violence, to make themselves masters of other men's persons, wives, children, and cattle; the second, to defend them; the third, for trifles, as a word, a smile, a different opinion, and any other sign of undervalue, either direct in their persons or by reflection in their kindred, their friends, their nation, their profession, or their name...Whatsoever therefore is consequent to a time of war, where every man is enemy to every man, [there is] continual fear, and danger of violent death; and the life of man, solitary, poor, nasty, brutish, and short.⁴

The problem is plain. We need to cooperate in order to establish some degree of stability. Yet, how in the state of nature can we establish the degree of assurance that is surely required in order for individuals to cooperate? How can we come to trust each other? Hobbes approach is instructive because he is not supposing that we take the state of nature seriously as a historical situation from which we emerged into a society organised through the institutions of the state. Rather, we are presented with a thought experiment whose aim to render vivid the problems we (as socialised moderns) would encounter should there be no state. Those problems arise out of our very nature, a nature explicated (to put matters crudely) in terms of the maximising rational agent model. No one can rely on agreements to mutual restraint of competition which they might make with others. For no one has any reason to assume that others will keep their word.

If a covenant be made, wherein neither of the parties perform now, but trust one another; in the condition of meer nature...upon any reasonable suspicion, it is void. (But if there be a common power set over them both, with right and force sufficient to compel performance, it is not void.) For he that performeth first, has no assurance the other will perform after, because the bonds of words are too weak to bridle men's ambition, avarice anger, and other passions, without the fear of some coercive power; which in the condition of meer nature, where all men are equal, and judges of the justness of their own fears, cannot be supposed. And therefore he that performeth first, does but betray himself to his enemy; contrary to the right (he can never abandon) of defending his life and means of living.⁵

The answer is that the establishment (or existence) of an effective sovereign power will put in place the potentially coercive framework that will enable individuals to rely on one another (for the most part): if you fail to keep to your agreements then the law provides the sanctions to punish you. I can rely on the self-interest of my counterpart and so come to the reasonable belief that she will keep to her word. The fact that there are sanctions constraining what it is in your interest to do cannot rule out the possibility that you will cheat. Assuming the sanctions are real and you are rational I can judge that you are likely to act reliably.

Now, one might object that there is more to trust than simply coming to a judgement as to whether someone else can be relied upon to keep their word. For on a simple reliance model of trust it seems that trusting someone becomes a function of assessing the probabilities. Thus trust would be:

a certain level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action.⁶

The difficulty in regarding trust as a belief about the reliability of others is in how to explain how we form such beliefs. In the case of the robber he has very good inductive grounds for relying on certain patterns of behaviour by victims, but this is in part precisely why that relationship does not look like the kind one can describe in terms of trust. Individuals trust one another under circumstances of uncertainty and risk. Trust is a response to a form of ignorance, an epistemological shortfall, that arises in certain circumstances. For example, imagine a Jew fleeing a Nazi death camp who comes to a church. He may trust a priest to assist him. There is clearly a risk in exposing himself to the goodwill of the priest for in the circumstances the escapee simply lacks the information that would enable him to come to a judgement on whether the priest is sympathetic to the Nazi regime or has other pressing reasons that would prevent him from offering help. Nonetheless, we surely want to say that trust can be placed in the priest because of *who* he is and what that entails. Of course, trusting someone cannot count as a guarantee of their performance. Trust has in this sense an asymmetrical quality. It flows from one person to the other, awaiting reciprocation but vulnerable to abuse. That *is* the risk. If I know enough about the situation to reliably assess the probability of someone performing, then I would appear not to need to trust them. I could simply rely on them. In a world in which there are psychics capable of reading minds one would surely be amazed to hear them talking of any need to trust other people. The puzzle is not how I can go about gathering enough information to do the sums working out the reliability of others, but what can justify or explain taking a risk on the goodwill of another when that information is not to hand.

Well, one might note that Hobbes has just the right kind answer. We cannot trust each other left to our own devices. However, establish a power to force us to act decently (this is not Hobbes' way of putting matters) and we have just the form of assurance required to overcome the reluctance to take a risk on others. The difficulty with this approach is that it may just be at odds with how we think about trust. This is not to say that we should just accept our countervailing intuitions, but we ought to explore

why an objection arises and whether we should stick with our intuitions or revise them.

The intuition I have in mind about our everyday understanding of trust is that it is distinct from mere reliance. The situations in which we employ the terms are not co-extensive. An account of trust just in terms of reliance may also appear to just collapse the distinction between reliance and trust. If to trust another is simply to rely on them, then there is an open ended range of ways in which I might properly judge them to be reliable in the circumstances in question. The grounds on which I determine the reliability of the other might include their stupidity. They can be relied upon to help me move flat because they cannot see how I regularly exploit their goodwill. Or, I might be able to rely on the cooperation of someone because I am her boss. I know she is ambitious and she will go to great lengths to impress me.

Consider once again the attitude of the fleeing Jew to the priest. We can hypothesise another situation of flight. A group of hardened criminals have escaped from prison. As they dash from the prison grounds they come to a wide and dangerous looking river, the howling guards and their dogs hot on their trail. Luckily the prisoners happen upon a robust rowing boat.⁷ In order to effect their escape they must co-operate in the rowing of the boat and in attempting to navigate the frightening currents of the river. As rational, self-interested individuals with a common goal each of them can rely on the others to do their bit. In our story it is clear to each of them that his own singular goal is only achievable through co-operating with the others, and each knows that each of the others knows this to be the case.⁸ We should say that the prisoners can rely on one another to do their part in escaping the guards. Once they have effected an escape, or at least one of them believes this to be the case, we have no reason to expect continued co-operation in the absence of some new external pressure. It is simply discordant with the way in which we usually apply the concept of trust to describe the situation as one in which the prisoners trust one another. Or is that simply to beg the question? Why ought we not talk of the prisoners standing in a relationship of trust?

The response to this challenge is to expose within an understanding of trust some feature that explains why the prisoners are not (obviously) trusting each other. The problem with the view that trust is reliance is that it fails to acknowledge that we trust others, and so see them as reliable, because we regard them in a certain way, and furthermore, in trusting them we are joined with them in a way of seeing the world. Our trust arises from our ways of seeing and interrelating with others. To see what we mean by trust we need to both generate cases where it appears to be salient to our understanding of what is going on, and to think hard about why.

First, trust may involve a willingness to be exposed to the risk of default by another because of who they are. Here one has not made any calculation about the likelihood of someone not acting in a certain way. Rather, the very thought of such a calculation does not enter into consideration. Following Annette Baier's way of putting matters trust is the willing acceptance of vulnerability to harm that others could inflict, but which we judge they will not in fact inflict.⁹ A trusting agent is one who has willingly accepted that his interests are exposed to the exercise of the goodwill of another.

A second feature of trust is that it involves an attitude or feeling towards another; it is (in part) constituted by an emotional state.¹⁰ To trust another is to have an attitude towards them that has a certain feel. The way trusting feels does not on the face of things appear to be like the way being in pain or despair is. Nonetheless, the way it feels to trust can be expressed in terms of the attitude of 'optimism that the goodwill and competence of another will extend over the domain of our interaction with her, together with the thought that the one trusted will be directly and favourably moved by the thought we are counting on her'.¹¹ That trust has a certain presence in our psychological state is also suggested by the sense of ill-ease, perhaps tension or even distress, experienced when one realises that trust is missing.

Trust is not just a matter of reliance because when we trust someone it is our attitude towards them that is essential in the belief or judgement we have with regard to their performance. Trust can be conceived as a complex of cognitive (belief-based) and affective (emotion-based) states. These states play an important role in securing cooperative behaviour in the face of risk, particularly in situations or contexts in which the extent, degree or probability of risk is (practically) incalculable. There is a further element in our understanding of trust: the idea that there is a non-instrumental good in trusting, which we may describe as a value that is *internal* to the practice of trusting. For I wish to propose that there is something that is good in trusting in *itself*. That is, there is a value in standing in relations of trust.

Thus far I have spoken in terms of one person trusting another, and played upon the sense in which trust is centrally characterised by the exposure to risk on the part of the trusting agent. Any elucidation of the concept of trust must attempt to cast light on the conditions in which some other can be regarded as trustworthy, and not simply adequately reliable in the circumstances.

Of course, by trusting each other we may be able to achieve through cooperative action that which would otherwise have been beyond us acting individually. Distinct from the good of the achievement of that goal is the value of being a certain way. That way is constituted by the very act of trusting and its component attitudes and beliefs. When a person trusts another – and that trust is not the subject of abuse - she comes to be assured that there is a commitment between them, a reciprocal bond upon which genuine or well-grounded trust rests and a shared understanding of the world presupposed in the act of trusting. If this is right, then by trusting we do not simply make possible the instrumental good of achieving shared ends or, indeed, of sustaining relations that are good in themselves such as friendships. Rather, by trusting we are engaging in a practice that is also in itself valuable and whose value inheres in the very relationship presupposed in the trust.

Care needs to be taken here in explaining just what is valuable about trusting. Trust can, of course, be abused. Sometimes the person investing trust in another is simply mistaken or foolish to do so. If in spite of all the evidence to the contrary I trust someone then I look to have erred. Trust is a relationship forged in the absence of reliance determining reasons, not one that can be sustained in the face of reasons that ought to compel me to judge that the other cannot or will not act as if he regarded my interests with goodwill.

Sometimes it will be a close call and at others not so. For example, if my friend with a compulsion for goldfish flesh offers to look after my fish while I am away, I am a fool to trust him. Indeed, if I care about the fate of the fish I am worse than a fool. Sometimes more than goldfish hang on our decisions on whether we ought to place trust another. In other cases we are not fools or naïve to place our trust in others. Certain of the relations in which we stand seem to presuppose trust in. All we need to know to take a risk on the other is who they are – instances of such trusted parties might include parents, teachers, priests and lovers. In virtue of the relations in which one stands to them there ought to be trust. That is, to be a friend or lover is partly constituted by bearing the strain of being an object of trust. If your parent, friend or lover asks why they ought to be trusted, then one may fear that they just do not grasp the proper nature of the relationship, or that the question signals the demise of the relationship.

When one person trusts another and they are justified in doing so the trust arises because they can understand each other in the appropriate fashion for the risk to be taken by the one and the faith or confidence being reciprocated in the actions of the other. Furthermore, when one is trusted the response typically extends beyond simply responding appropriately in the sense that trust does not spring forth from a vacuum. To trust (and be trusted) is part of the typically complex array of practices and attitudes through which the domain of social interaction is constituted. On this view trust is to be understood as a way of regarding others which is embedded in the networks of social practices. Widespread trust is possible because it is presupposed in our ways of going on together. That capacity to trust is as secure or as vulnerable as those practices through which relations, institutions and groups are sustained, and in particular the capacity of those practices to withstand the pressures of critical reflection on the part of those engaged in them and the challenges of other practices.

Now, to return to the question of the value of trust. The value in trust is the value we enjoy through standing in a trusting relationship. Moreover, the value in trust is objective in the sense that the good is not a question of how I feel. Rather, the relationship both forged by and constituting the trust between us, is the source of a good in virtue of which our lives go (at least potentially) well. Trust is presupposed in the nature of certain relations; that is, trust is part of what it is to stand with another(s) in a relationship of love or friendship. In standing in these kinds of relations each of our lives is enriched.

My observations on the meaning of trust do not amount to an argument that this is how we should indeed understand that term. Rather, they represent an attempt to analyse how we can think about a concept, and as such can only represent a fragment of an endeavour to examine the nature of trust. That enterprise might need to abandon the attempt to provide a univocal analysis for the idea of trust may simply not be a settled term with a single application across different domains of discourse. I shall conclude the paper by briefly assessing if the analysis of trust offered meets certain bare requirements one might expect from an account of trust. According to Jones ‘an adequate account of trust should be able to explain at least the following three fairly obvious facts about trust: that trust and distrust are contraries but not contradictories, that trust cannot be willed, and that trust can give rise to beliefs that are abnormally resistant to evidence’.¹²

The account given accepts that at a general level distrust is a contrary of trust. Bob and Jane may just have no attitude or distrust towards each other in the relevant context. The failure to trust is not necessarily equivalent to distrusting, standing in a relation of pessimism about the goodwill and motivation of the other. In some contexts, though, the failure or inability to trust may signal distrust. If I do not trust my wife when she sets out her plans for the day then it can not be explained as lacking a particular attitude towards her, as being locked in 'neutral'. The thicker or deeper the relations in which we stand then, perhaps, the more compressed becomes the space between trust and distrust.

That trust can not be willed seems right and supported by the account given. Emerging from our interactions, to trust means that one stands with others and sees the social world (in the appropriate context) in a distinctive way. I might need to decide whether the grounds for trust are really present, I might reflect on my feelings towards the other(s). Such considerations impact upon the stance I take towards the world. However, I merely go wrong if I begin to perceive the world as if I do trust certain others; that is, if I *decide* to act as though I enjoyed the perspective of one who stands in certain trusting relationships. To simply decide – to, as it were, declare to myself - that I trust someone may induce in me the beliefs and feelings associated with trust. However, their source is not located in a relationship with the other person, and those beliefs and emotions are to be regarded as counterfeit. This need not mean that trusting relations are therefore personal in being proximate or close in social space or face-to-face. For I can trust those who I have never met through the mediating chains of relations and practices. So I may trust the leaders of my faith or community. However, to trust in a way that is appropriately grounded in the ways in which I am connected or related to the other I need to be so joined. It is not simply down to my act of will.

When individuals trust one another there is a resistance to accept evidence that undermines that trust. Again this is conspicuous where trust is deep and where it is often associated with other affective attitudes. Bob may soon learn that he is a cuckold. Indeed, the evidence of his wife's infidelity has been mounting. For an outside observer it appears nearly impossible that he could have missed so much that was so obvious. Yet, Bob trusts his wife because he loves her. Because of his understanding of the relationship in which they stand – or more accurately, the relationship in which they once stood and which he believes they continue to stand in - he interprets evidence of his wife's adultery into innocence. When we trust coherence is brought to part of the world. That coherent form serves as a constraint on how we respond and act and as a filter through which evidence is interpreted. Trust is sometimes difficult to undermine because we do not merely have a desire for coherence and stability. It is rather that the form or structure of our world is determined in large part by the nature of those relations in which trust is so central. A world that is dear to us, familiar and one that constrains our understanding is itself threatened by the dissolution of trust. Perhaps our resistance to the evidence against trusting resides in the living of a kind of life.

Paul Sheehy
Richmond-upon-Thames College

¹ An oversimplifying account of this nature applied in direct empirical enquiry is doomed to misconstrue every individual act as being motivated either by the play of short-term consideration or as interpreting every motivation as being directly connected to the agent's interests. Things are likely to be considerably more complex. For example, not all social structures, practices and attitudes need be explained in terms of an agent's awareness of self-interest. It may be the case that they can arise and be sustained because they are functional for the self-interest of agents individually or for the group as whole. Alternatively, there may be *no* interest promoted by such attitudes or practices, but they do not serve to undermine or compete with such interests. The present point is that the rational maximising agent concerned with her self-interest has featured as the main character in the ways in which social interaction has been modelled in philosophy and the social sciences. The figure of the 'economic man', *homo economicus*, has loomed large in economics. This paper neither endorses nor presupposes that view of human nature, but suggests that such a view prompts a puzzlement about trust. Of course, there are alternative views about human nature. For example, one might hold that we are disposed to co-operate and be other regarding. In which case one might wonder why trust is accorded such importance.

² Philosophical enquiry in general is interested in the concepts we employ in our ways of going on – in what we say. We are particularly interested in those concepts which apply with the broadest scope so that they frame or delineate the subject matter at hand. A cluster of apparently unrelated statements are united by their dependence on a single concept. Taking an example from metaphysics one might note that Paul is short, the dog happy and the housing market overpriced. In each of these statements a claim is made that some entity possesses a particular property. The concept of *property* is central to an understanding of what is being asserted. In metaphysics we can also include, *inter alia*, cause, space, time, entity, substance, identity in a list of basic or 'domain concepts'. See Thomas, G. (1993) *An Introduction to Ethics*, London: Duckworth pp. 2-4 for an excellent discussion of conceptual analysis and its role in ethics. A rewarding and more advanced discussion of conceptual analysis is Frank Jackson's (1998) *From Metaphysics to Ethics*, Oxford: Clarendon Press.

³ *Collins Third Edition* (1991). It is thought that trust is derived from the Old Norse *traust*, which means confidence or firmness. This has been associated with the Indo-European root *deru* or *dreu* meaning to be firm or solid.

⁴ Hobbes, T., *Leviathan* Chapter 13.

⁵ *Ibid.* Chapter 14.

⁶ Gambetta, D. (1988) 'Can We Trust Trust?' in D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*, New York: Blackwell pp. 213–37.

⁷ The story of the escaping prisoners is taken from Hume's example of two men rowing a boat in his analysis of conventions. See Hume, D., (1740/1978) *A Treatise of Human Nature* 2nd Edition (ed. Nidditch), Oxford: Oxford University Press p. 490.

⁸ For example this might become abundantly clear in their conversation and actions upon arriving at the river.

⁹ See for example, Baier, A., (1994) *Moral Prejudices*, Cambridge, Mass: Harvard University Press p.152.

¹⁰ I leave to one side any further discussion of how we should properly understand an emotion, a topic which has attracted a considerable amount of attention in recent years.

¹¹ Jones, K., (1996) 'Trust as an Affective Attitude', *Ethics* 107.

¹² Jones op cit p15. We call terms (or propositions) contradictory if they cannot both be true and they cannot both be false. So, being alive and being dead cannot be true of anything at the very same time. If I am alive it is impossible for me to be dead, and it is impossible for me to be neither alive nor dead. Terms are contraries if they cannot both be true, but can both be false. Thus I cannot both love and hate you, but it is possible to neither love nor hate you. I may just be indifferent to you. So with trust, I may not trust you but that does not mean I distrust you. I may just have no attitude towards you at all as far as trust is concerned.