WHAT IS TRUST?

BY

THOMAS W. SIMPSON

Abstract: Trust is difficult to define. Instead of doing so, I propose that the best way to understand the concept is through a genealogical account. I show how a root notion of trust arises out of some basic features of what it is for humans to live socially, in which we rely on others to act cooperatively. I explore how this concept acquires resonances of hope and threat, and how we analogically apply this in related but different contexts. The genealogical account explains both why the notion has such value for us and why it is difficult to define.

1. Introduction

Here are two observations on the concept of trust. First, I suggest it has a trumping effect. When the term 'trust' is deployed in conversation, it has a way of relegating other values to a lesser status. Think of two business partners discussing their relationship; it is one thing to talk about ways to improve communication in the workplace, but quite another to talk about ways to restore trust. Once questions about trust have been raised, other concerns tend to be crowded out. Second, however, trust tends not to be talked about very much. Most of the time, it is an invisible assumption. By way of tacit confirmation, the philosophical literature on trust remains slim indeed, especially when compared to the attention given to concepts of comparable import, such as justice, knowledge or truth. These two observations do not sit comfortably, however, as important concepts tend to get talked about.

The task of this article is to elucidate the concept of trust. The paradigm way that analytic philosophers have addressed tasks of this nature is through conceptual analysis, assuming that the rules of correct use are determinate and therefore statable, perhaps because there is some particular phenomenon to which the term refers. At its zenith, the project issues in a definition stated in necessary and sufficient conditions, using only terms understood more clearly than the definiendum. The influential philosophical treatments of trust have generally assumed something akin to this: 'trust is *this*'.

Pacific Philosophical Quarterly 93 (2012) 550–569 DOI: 10.1111/j.1468-0114.2012.01438.x

This seems to me the wrong way to go. There is a strong *prima facie* case for supposing that there is no single phenomenon that 'trust' refers to, nor that our folk concept has determinate rules of use. Nonetheless, this does not mean that there is no philosophical understanding of the concept to be had. In this article I develop a genealogical account, approaching the task by asking: what human needs does the concept of trust answer to?

The answer I propose is that in living socially, people must rely on others to act cooperatively. So it is highly desirable for any community of language users, who necessarily live socially, to possess a simple root concept to refer to this fact. Trust initially answers to this. But particular contexts of use in which this simple root is salient lead to it acquiring richer and not easily definable resonances, both sentimental and normative, regarding hope for sociality and the threat of conflict. Thus the concept of trust acquires nuances additional to its root. As well as acquiring these resonances, I argue that particular features of this concept make it apt for use in contexts where there is a strong analogy to its originating conditions. These are sufficiently recurrent, and the use of 'trust' so apt in them, that we must recognise plural forms of trust, reflecting extensions and developments of the central notion. Over rival elucidations of trust which give an analysis of the concept, the genealogical account has the significant merit that it fits the facts. Not only so, it has the further merit that it explains why trust has such value for us, the point which underlies the trumping observation, yet also why it is so often an invisible assumption.

The structure of the article is as follows. I first make the case that trust is not amenable to conceptual analysis, and outline the merits of a genealogical approach (§2). I then develop the account itself. Social living constitutes the generative conditions in which a simple root concept arises, one that refers to the fact of reliance upon cooperation (§3). The contexts in which 'trust' is generally used lead to the term acquiring rhetorical potency, and thus the concept itself acquires nuances of meaning (§4). The term is then used in contexts different but analogically related to the generative conditions. Its use in these contexts is so apposite that, by repetition, variegated derivative forms of trust arise. Thus 'trust' is an umbrella term under which plural notions cluster (§5). I conclude by noting the implications of my argument for enquiries that both make use of or take as their subject the concept of trust, using Paul Faulkner's recent work with trust in the epistemology of testimony as a test case (§6).

2. *Analysis and genealogy*

Why should trust not be amenable to conceptual analysis? Against some, I do not think that conceptual analysis is always a hopeless project. Nor do I have an argument to show that trust *must* be too heterogeneous to be

defined, so I have no knock-down reply to the question. But even if no conclusive reasons are available, I suggest that some forceful ones can be given for not persevering with attempts to analyse trust, and for trying a different way.

An inductive reason is the ease with which counter-examples can be produced to existing definitions. Three major ways of thinking about trust have emerged in the literature. The first significant philosophical treatment was Annette Baier's, who endorses a broadly affective approach. She claims that 'When I trust another, I depend on her goodwill toward me' (Baier, 1994, p. 99). Instances of trust abound, however, where goodwill does not seem to have much to do with it. Against Baier, Onora O'Neill makes the point nicely that a patient may trust a doctor to exercise proper professional judgement in their case, while knowing full well that the doctor finds them particularly irritating and bears them no goodwill (O'Neill, 2002, p. 14). Not only is goodwill not necessary for trust, it is also not sufficient. Richard Holton points out that a conman may rely on their victim's goodwill, while not trusting them (Holton, 1994, p. 65).

While Karen Jones concurs with Baier that trust has a distinctively affective dimension, she adds a role for affect to the trustor as well as the trusted. Trust 'is an attitude of optimism that the goodwill and competence of another will extend to cover the domain of our interaction with her, together with the expectation that the one trusted will be directly and favourably moved by the thought that we are counting on her' (Jones, 1996, p. 4). As well as O'Neill's objection, this is vulnerable to further counter-examples where trust seems to be a matter of cold calculation on the part of both parties. Think of a group of oligarchs arranging to price-fix. There is no love lost between the ruthless competitors. But they may still successfully manage to collude on raising prices over a staggered period of time—to avoid the suspicion of coordinated action – with at least the initiating party having to trust that the others will follow her lead and not take advantage of the price differential to increase market share. The initiating oligarch does so only because she has good reasons to believe that the others will follow; optimism does not come into it. Trust is not always an affective matter.

Richard Holton's treatment of trust is equally influential. Rather than a primarily affective relation, his account sees trust as characterised more by normatively-laden attitudes. 'When you trust someone to do something, you rely on them to do it, and you regard that reliance in a certain way: you have a readiness to feel betrayal should it be disappointed, and gratitude should it be upheld' (1994, p. 67). This range of reactive sentiments is a result of adopting 'the participant stance' (the idea is adopted from Strawson, 1974). Again, it is not hard to find cases of trust that do not fit the description. A mother certainly could be relied on by her son while he adopts the participant stance towards her. But it is odd to suppose

that the participant stance is doing any distinctive work here. The son trusts his mother because he knows she loves him. He need not even have the concepts required to adopt attitudes such as those mandated by the participant stance. Philip Nickel's closely related account declares trust always to involve the ascription of an obligation to another (Nickel, 2007). It falls foul of the same kind of counter-example.

A different way of thinking about trust sees it as fundamentally a matter of rationality, understood with the economists as forward-looking, interest-maximising action. The baldest, most austere version is James Coleman's. He asserts that 'the elements confronting the potential trustor are nothing more or less than the considerations a rational actor applies in deciding whether to place a bet' (1990, p. 99). One 'places trust' if one acts in a way that relies on another person, and which thereby exposes the trustor to risk. But defining trust as an action, which one places or not, fails to describe situations where trust is latent. Out of kindness to the child she never had, a rich aunt promises her nephew that if he ever falls on tough times, she will bail him out. As it happens, the nephew does well and wants for nothing. Yet he still trusts that she would have fulfilled that promise if things had not gone so well. So the nephew trusts, but never acts in a way that relies on another.

Russell Hardin's encapsulated interest account of trust is similarly inspired by rational choice models of action. He defines trust as simply a belief about another's trustworthiness. 'The declarations "I believe you are trustworthy" and "I trust you" are equivalent' (2002, p. 10). Belief that the other is trustworthy, though, is hardly necessary for trust. A father may give his daughter some money and a list of shopping to get from the local store, despite her known predilection for cola bottles. He may do so as a form of moral training, despite having no fixed belief that she will be trustworthy, or even despite believing that she will succumb to the temptation. There is surely a permissible sense in which he trusts her.

As reasons of space prohibit surveying all claims of the form 'trust is *this*', I can do no more here than report that counter-examples can be similarly easily generated to all those I have found. I suggest the reader will find the same. As the above authors represent the three major ways of thinking about trust in the literature, this is inductive reason to think that other such attempts will be similarly vulnerable to counter-example. As noted, however, trust has received relatively little attention from philosophers, especially when compared to endeavours like the analysis of the concept of knowledge. So a defender of the possibility of an analysis of trust might reply that the fact that existing definitions are vulnerable to counter-example does not show that trust is not amenable to such a treatment. Perhaps no one has given it a really decent shot yet.

I am dubious. The ways that the word is used are simply too various to be regimented into one definition. Sometimes 'trust' is naturally

understood as referring to a sort of affective attitude ('I will trust my husband, I will not be jealous'); at other times to a conative one ('Come what may, I will trust you to the end'); and at yet others to cognitive ones ('I know you are an honourable woman, so I trust you'). Sometimes it is not a mental state but action which is described as trust ('The patrol followed the scout, trusting him to spot any ambush'). Similarly, it is used in situations where the motivation to trustworthiness is dramatically varied: love, or mutual gain, or moral considerations may all count as reasons not to betray someone's trust. These all support the inductive argument against the plausibility of analysing of trust. Counter-examples can be given so easily *because* there are so many ways the word may permissibly be used, and so it would be foolish to seek a single definition.

This observation is not yet decisive. For a defender of the possibility of conceptual analysis may propose a disjunctive analysis: 'You trust someone if and only if *either* you rely on their goodwill *or* you adopt the participant reactive stance towards them *or* you believe it is in their interests to be trustworthy *or* . . .'. The advantage of the strategy is that any plausible view about trust can be accommodated, and so the analysis can be made immune to counter-example. The disadvantage of the strategy is that any plausible view about trust can be accommodated. It can be made immune to counter-example not because it is true but because it is ad hoc.

Arguably the most significant defence of the possibility and necessity of conceptual analysis is that given by Frank Jackson (1998, 2005). Jackson points out that, for most of the concepts which philosophers are interested in, there are supervenience relations with other concepts; pretty much everyone agrees that knowledge supervenes on truth and belief, for instance. But if so, he asks, how can knowledge be *sui generis*? I accept the point. Yet parallel claims apply to trust, for we make judgments about whether someone trusts another in virtue of more basic features of that relation, such as whether they are relying on them, and what the reasons for their reliance are. So as trust supervenes on other more basic concepts, I do not claim that it is *sui generis* and unanalysable. But if trust is not unanalysable, why not analyse it? Jackson himself provides the answer. Granted that most philosophically interesting concepts are not *sui generis*, he still needs to explain why conceptual analysis is so hard. Part of his explanation is a recognition that different people can use the same word to represent the world in subtly different ways, with representation for Jackson a matter of referring to patterns in the world. He concludes that 'smart philosophers' recherché examples . . . reveal that there are a number of candidate patterns [referred to by the same word]. What the counter-exampler refutes is the view that there is a single, fixed concept which we all, or nearly all, use the word "knowledge" for' (2005, p. 135).

The vulnerability to counter-example of putative analyses of trust warrants the same conclusion. There is no single, fixed concept which we all, or nearly all, use for the word 'trust'. Your disagreement with my analysis simply reveals that we represent the world differently with that word. There may be no fact of the matter about which is right.

As well as the above case for the unsuitability of trust for conceptual analysis, there is further reason for trying a different approach. Even if a successful analysis was given, there are some interesting itches that such a result would fail to scratch. In particular, why should trust seem so important a concept in our practical lives? Simply establishing that a concept answers to a particular set of conditions holds little prospect of answering that.

An alternative way of thinking takes the value of trust as of first importance. A genealogical approach addresses the concept obliquely, by asking why we might have the concept that we have, given some broad facts about how we live and the projects we pursue. A full defence of this method is beyond my scope here. I am indebted for it to the fertile treatments of knowledge by Edward Craig and truthfulness by Bernard Williams, and must refer the unconvinced to their longer discussions of the strategy, from which the following brief remarks are drawn.[1]

The claim is that if we can give a 'role description' for an important concept that looks and feels very similar to a notion that we actually operate, explaining what is needed of a concept in order to do a particular job, then that sheds light on the content of our actual one. Rather than trial by thought-experiment, it instead accommodates the vagueness and conflicting intuitions that surround difficult and abstract notions. No doubt all the advocates of the 'trust is *this*' claims I critiqued above have replies to my counter-examples. But the sorts of replies on offer will tend to consist in argument over the cases, and ultimately a discovery that we use the word 'trust' to describe different things on different occasions, illustrating Jackson's point above. What would be genuinely revealing is an explanation of why the term permits this variability. As well as yielding a clearer grasp on the content of the concept, such an approach also makes it entirely perspicuous where its value comes from. Two results for the price of one enquiry is no bad thing in these straitened times. Finally, the genealogical method is broadly naturalistic in that, with Williams, it 'helps to explain a concept or value or institution by showing ways in which it could have come about in a simplified environment containing certain kinds of human interests or capacities' (2002, p. 21). Although an analysis of trust cannot be ruled out of court, the burden of proof remains squarely on those who espouse its possibility, and alternative approaches should be seriously entertained. As to the possibility and fertility of a genealogical account – to which I now turn to develop—let the proof be in the pudding.[2]

3. The generative conditions

Although it is a platitude that man is a social animal, it is an important platitude, and it is the starting point for reflection on the concept of trust. The isolated life, when there is no possibility for sociable contact, has a quality of the nightmare about it. Robinson Crusoe *endures* his isolation until Man Friday arrives, questioning what God intends by providentially requiring this ordeal of him, and so illustrating quite how deeply abnormal it is to live alone (Defoe, 1719, p. 107). Yet even Crusoe had a father and mother whom he was aware of hurting when he ran off to be a sailor, and the idea that a significant population of *homo sapiens* might each live in conditions of thorough-going isolation, except for the bare necessities of mating and parenting, is so *outré* as to forbid that population being described as human. Social contact constitutes a very great share of the well-lived life, in which we laugh, are musical, gossip, adorn the houses we share, and otherwise enjoy passing the time together. In describing ways of living and acting that are possible only if humans live socially, I do not suggest that it is somehow a voluntary decision to undertake them.

Yet the nature of the projects that are possible only if humans live together is significant. Perhaps the kinds of projects that are most obviously made possible are those that require the joint action of several individuals towards the same end. Crusoe was able to fell a tree and carve a dugout canoe from it by himself, but he only discovered later, and to his cost, that the craft was too heavy for him to push to the sea, so he had to abandon it in the forest (1719, p. 137). It takes little imagination to multiply other examples of joint action from evolutionary just-so scenarios, of groups of humans being able to hunt down prey or defend themselves against predators that would easily evade or overcome a lone individual.

The examples above of hunting and protecting are reassuringly simple instances of joint action, for they portray a situation where you and I must cooperate together for this same period of time, and then, if successful, we both enjoy the same benefit at the same time, whether it is relief from danger or being fed. But the Crusoe example opens up the possibility of a more complex form. It need not be the case that you and I both work on the dugout canoe at exactly the same time if we are to make a canoe and both benefit by it. Individual action can be separated over time and still successfully be joint. Separating participation by time is more complex because it demands that both parties are able to take account of the other not only in their presence, but also in their absence, which demands a degree of socialisation.

As well as joint actions, however, there can be social forms of action with individuals acting in a coordinated way, but nonetheless towards different ends, ends which would not be achievable, or at least not so

readily, by one person acting alone. I may be particularly good at making pots, and you particularly good at making ploughs. So I give you a pot in return for one of your ploughs, and we both benefit. Joint actions and exchanges are both instances of mutualism, where both of us benefit in ways that would not be possible without cooperation. Like joint actions, exchanges can also be separated by time. I might offer you some fish I have caught now, in return for your giving me some wheat later this autumn. But exchanges separated by time have an additional complication, in that for one party at least there is a more beneficial option: having got the fish, you now might decline to give me the wheat. Some joint actions are subject to the same temptation; letting others go and man the barricades means I do not run the risk of death. These situations are more socially demanding, because they require people to recognise the collective benefit of not doing what is in their maximal individual interest.

A further form of coordinated action is also necessary for social life. The examples so far discussed are well characterised as 'positive' projects, with people trying to make things better for themselves and others. Equally essential are 'negative' projects, where people must not make things worse for others. Bringing rapine, violence, and other anti-social behaviour to a sufficiently low level for those who live in a community is a basic precondition of the well-lived human life; life is still possible in an environment of constant threat, but almost everyone finds it very unpleasant. It is not necessary that everyone without exception abstains from such acts; all that is necessary is that enough people act cooperatively to secure a relatively threat-free social environment.

I suggest that these various forms of shared existence—the domestic life of child-rearing and shared company, of exchange, and of joint, positive and negative collective action—constitute the basic forms of social life. I also suggest that these provide the generative conditions in which the root notion of trust arises. Call this simple root *Ur-trust*. Three features are salient.

First, these social forms of action are all occasions when people rely on others. Although the notion of reliance is commonsensical, it has an important implication worth highlighting. For reliance permits of degrees. You are very reliant on me if the consequences of my unreliability are very serious for you. You can also be reliant on me to a greater or lesser degree in achieving a particular goal. These two dimensions of variation are separable: I may be essential to your achieving some trivial goal, or I may be tangentially involved in securing a very important goal. Because you rely, you are exposed to risk of loss.

The second feature of these social forms of action is that they require people to behave cooperatively. The notion of cooperative behaviour can again be left largely intuitive, although I take it that it should not be reduced to rule-observance, but rather is best stated in terms of motives to

action. To behave cooperatively means that you take account of others' good in how you act. At the least, this means ensuring that your action is not directly detrimental to others. It may also mean acting for their good when appropriate.

How costly is cooperative behaviour? Sometimes, it is not very costly at all. Negative collective action merely requires abstaining from possible gain, and that can often only be won through some act of violence with its concomitant risks. Even here, though, gain foregone through being negatively cooperative may nonetheless be significant. Other forms of cooperative action may be more costly. Keeping a promise, sometimes, can be a demanding thing to do, if things have turned out differently to the way you expected when you made it, or you made the promise rashly, without considering its implications. Yet breaking promises is a very uncooperative thing to do. Other instances of collective action can also be personally costly; sometimes being cooperative involves refusing the temptation to free-ride and paying your share of the bill, even though you could get away with not doing so. Cooperative behaviour is most valuable and in most jeopardy precisely when it requires overcoming the temptation to personal gain, either foregoing gain which would have been at others' expense or contributing your part towards the shared project.

A third feature of these forms of social action is that, as well as a risk of loss, there is uncertainty as to whether those I interact with will be cooperative. This is for the simple reason that the workings of others' agency are not under my control—people are free to decide what they do. (No metaphysical commitments are made; the compatibilist as much as the libertarian can endorse the point.) I can influence others' decisions, to be sure. But if I were able to control what someone else did, then it would no longer be her who was acting, and so it would not be *her* cooperation that I was concerned about.

In a social world, then, where we raise families and enjoy living alongside each other, where others let us get on with what we are doing when it does not affect them, and where we can achieve things together that we could not do alone, I am routinely reliant on others' free and cooperative action. This, I propose, constitutes the Ur-notion of trust: I trust someone when I rely on their freely cooperative behaviour. This reliance is a property of my action, in the first instance. The importance of reliance on others' cooperative behaviour to the continued existence of a community makes it highly desirable that they possess a word in their language to describe that property of action. The term 'trust' enables English speakers to refer to the fact of reliance on others' cooperative behaviour (*mutatis mutandis* for other languages); this is what the term is for. On this Ur-notion of trust, it is not that action reliant on others' cooperative behaviour is evidence of a preceding or contemporaneous mental state which provides part of the practically rational basis for that action, and which we call 'trust'. Rather,

that action *is* trust. An implication of this is that Ur-trust is possible for those who act without conscious deliberation, but in a way that is reliant on others' cooperative behaviour, such as infants or the mentally impaired. This seems correct, or we would not take infants' thoroughgoing dependence on their parents as a paradigm kind of trust.

The root notion says nothing about why people act cooperatively. All that matters for Ur-trust to arise is that people do rely on others to act cooperatively often enough to need to talk about it. For this reliance to be a stable pattern of behaviour, people must actually be cooperative relatively frequently. But there need be no particular reason why they are cooperative.

The Ur-notion provides a clear explanation of why trust comes in degrees. The variability in the cost of unreliability is one source of the variability in degree of trust. If another person were to act uncooperatively in a way that resulted in very serious harm for me or my interests, then my reliance is correspondingly greater. The degree to which I rely on another person in pursuing my projects constitutes another source of variability. And there is a third source of variability, the degree of uncertainty about whether another person will prove cooperative. If cooperative behaviour does not cost much, for instance, then there is less chance of failure, and thus less trust is required. But if cooperative behaviour would be more costly, then the likelihood of failure is higher, and thus more trust is exhibited if the truster chooses to rely. Soldiering represents the limits of this, where all three variables are pushed to or towards the maximum: your life is at stake; you are nearly wholly dependent on others playing their part for you to stay alive; and because the cost to others of taking account of you may well be their own life, leading to great temptation not to be cooperative, so there is considerable uncertainty. In these respects, soldiering is a feat of trust. (Rock climbing pushes the first two variables to their limit too; it too is a feat of trust.) So the Ur-notion of trust, of reliance on freely cooperative behaviour, comes in degrees.

The type of action referred to by the root notion also explains why trust should be so valuable. As social existence would not be possible without reliance upon cooperative action, so threats to Ur-trust constitute threats to the continued existence of society. The analogies here become superlative: trust is like the air we breathe (Baier, 1994, p. 98), or the cement that holds society together (Acton, 1974, p. 14). The genealogical account explains why such analogies should be pertinent.

4. The rhetoric of trust

The Ur-notion of reliance upon cooperative behaviour does not exhaust the complexity of the concept of trust. I suggest that certain contexts of use

of this root create added resonances, which amplify and subtly alter the concept itself, towards the richer notion that we actually apply. It is this that accounts for both the invisibility of trust, and for its rhetorical force when it becomes the subject of conversation.

As a starting observation, it is a basic datum of the psychology of action that a very great proportion of what we do is not the outcome of a formal process of deliberative, conscious reasoning. (This implies nothing about the rationality or otherwise of that action, in any but a trivial sense of that protean term.) This is no less true for cooperative action than it is for individual action. Habit, not ratiocination, is the standard mode in which we conduct ourselves, although we switch to the latter in times of stress or uncertainty.

This observation, however, has a striking consequence for the concept of trust. Because reliance on others to act cooperatively is such a routine part of life, we very often trust without talking about it. Most of the time, it just happens. The actuality of trust may be very present; but it does not need to be talked unless there is some problem, and so trust is invisible. So it is precisely non-routine contexts that generate the need to start talking about trust. 'Non-routine' is too bland a description, however. It is contexts which are new and unfamiliar, or in which things have gone wrong, or where there is some particular reason to worry that things might wrong—perhaps where the stakes are particularly high so that I am very reliant on others—that there is pressing need to draw attention to the fact that I am relying on someone else to act cooperatively. It is in the breach that the term 'trust' is particularly apposite. As such, it acquires a resonance of crisis. Talk about trust functions as an alarm bell; when it goes off, it acts as a signal that, for some reason, the habitual assumption of cooperative behaviour no longer applies. The Ur-notion of trust as simple reliance on cooperative behaviour thus acquires a rhetorical resonance, as a warning signal to stop and think about what you are assuming of others. The very fact that someone has asked the question 'Should I trust?' comes to imply that there is a problem in the offing. Paradoxically, talk of trust can be a potent indication of suspicion. The fact that talk of trust can be an indicator of suspicion has a further corollary, that it is only when we *stop* talking about trust that the kind of trust sought as the precondition of social living can be presumed.

'The breach' is a metaphor. This should not be allowed to obscure the severity of situations where the assumption of cooperative behaviour no longer applies. Widespread anomie and social breakdown are recurring features of dystopian visions of the future, or of warning stories from the past. Deep and irreversible conflict has very significant imaginative potency. The fact that anomie is so serious explains why trust—in invoking its spectre—tends to trump other values in conversation.

The role of the breach as a prominent context of use of 'trust' is not the only source of additional resonance, though. The Ur-notion of trust says nothing about why people might act cooperatively. But consider what David Hume called the 'principle of humanity', consisting of a basic 'feeling for the happiness of mankind, and a resentment of their misery', which he supposed was nearly universally shared, although to varying degrees (1748, pp. 272, 286). Those who have this basic feeling in even some small respect cannot but regard cooperative behaviour with approval, and resent uncooperative behaviour. If people are willing to act on this basic feeling, then they will express approval of and cooperate with those whom they see to act cooperatively, and express disapproval of, avoid and sanction those whom they see to act uncooperatively. In a world where most people are motivated by the principle of humanity to a non-trivial degree, it will generally be prudentially wise to act cooperatively. This is the case even for Hume's 'sensible knave' who has no fellow-feeling for mankind (1748, p. 282). So drawing attention to the fact or otherwise of trust cannot but carry with it the promise of approval and goodwill towards the cooperative, and the threat of disapprobation, hostility and sanctions towards the uncooperative.

The same promise and threat, however, can derive from an at least arguably different source of motivation other than sentiments of goodwill to humanity. A concern for justice may be equally effective in motivating cooperation. Although I presume that cooperation is not itself a moral notion, it is clear that there are times when I am obliged to be cooperative. If I am entitled to require some form of cooperative behaviour from someone else, then my trust comes with a normative expectation that they ought to do the right thing. In these contexts, the community's requirement that one follow its injunctions means that trust rides intimately alongside the full force of the moral system's demand for compliance.

Less common, but nonetheless possible, are situations where being cooperative would be a supererogatory act worthy of praise. I may rely on someone else to take me into account in a way that goes well beyond what may be expected of them. If they decline to go the extra mile, more fool me for acting on mere hope. But if they come up trumps, then trust is here an heroic testament to the possibility of human self-sacrifice for others, and an emblem of the best of what we are capable of. Trust is sometimes associated with hope for the possibility of unexpectedly and heroically compassionate action, favouring others over oneself.

'Trust', then, comes to acquire other resonances. One is of hope for the possibility of the broad sunlit uplands that beckon if people could live together harmoniously, being kind and considerate to one another. Another resonance is of the threat of interpersonal conflict: screw me around and I'll make you pay, and hopefully the community will too. Think what it is to be asked, 'Can I trust you?'. It is impossible to answer

'No' without near ending a cooperative relationship, while answering 'Yes' is a tacit acceptance of the questioner's ill-will and possible revenge should you not come up with the goods. It is not a kind question to be asked at all.

The latter resonance—of threat—sheds light on a related feature of the way in which our practices of trust and the associated language shape interpersonal relations. If it is explicit in the way that I act that I have refused to allow myself to be reliant on your cooperation, or I straight out tell you that I do not trust you, then something has gone seriously wrong between us. For in showing that I believe it would be foolish for me to assume that you will be cooperative, I declare to you and to all who learn of my refusal to trust that you are the sort of person who is out only for themselves. It is an insult to refuse to trust. The fact that refusing to trust can constitute a social rupture is indubitable, but the point needs qualification. Sometimes the stakes are so high, or the temptation to be uncooperative so significant, that everyone recognises it is no offense to decline to rely fully on another's cooperative behaviour. If this were not the case, insurance policies to cover others' malicious behaviour would always be immoral, and that seems unlikely. Relatedly, there is a long tradition of thought which regards government in this light, seeing it as proper to distrust all public officials due to the temptations of their position to abuse power. Distrust is also the proper attitude in domains of life where competition and conflict are approved.[3]

A further feature of trust is not well described as a rhetorical 'resonance', but arises out of those just noted. Hume was not the only sentimentalist to note the significance of a basic feature of human psychology for the moral dimension of our interpersonal relations. Adam Smith observed another.

> Nature, when she formed man for society, endowed him with an original desire to please, and an original aversion to offend his brethren. She taught him to feel pleasure in their favourable, and pain in their unfavourable regard. She rendered their approbation most flattering and most agreeable to him for its own sake; and their disapprobation most mortifying and most offensive (Smith 1790, p. 135).

Smith's claim need not be taken as universally, as true of *all* people. But as a broad generalisation, it is surely correct that most people find pleasure and satisfaction in being liked and esteemed by others, and the reverse is correspondingly painful. When this point is conjoined with Hume's observation about the principle of humanity, though, a further interesting consequence arises. For if I can broadly assume that others will approve of my cooperative behaviour, and I also find pleasure and satisfaction in that, then this provides with a reason to be cooperative. Moreover, this operates regardless of either the collective benefit that cooperative action will achieve, or any long-term personal benefit I will gain from the reciprocal

cooperation of either that specific person or generalised reciprocators who are aware of my cooperative track record. The fact that I trust another is a direct encouragement to the one whom I trust to act cooperatively. Even more is this the case when the fact of reliance is directly referred to by invoking the term 'trust'. So there is a further nuance in the notion of trust: it is a bootstrapping attitude. The very fact of publicly avowing my trust helps to create conditions that justify it.[4]

The foregoing needs to be more carefully stated, however, as Smith's point needs to be qualified. It is a relatively rare individual whose desire for esteem is indiscriminate; for most, there are specific people or groups whose esteem is valued, because they respect them particularly. Indeed, I may actually take it as a compliment if those whom I have no time for in turn *dislike* and disapprove of me—at least they noticed me. So publicly avowing my trust will not always succeed in bootstrapping. Rather, it is the fact of publicly avowing my trust to those who are likely to value my esteem, or to those who are likely to value the esteem of those who they know are watching and think that I ought to be cooperated with, which helps to create the conditions that justify it.

The observation that trust is, at least on occasion, a bootstrapping attitude, also illuminates a further point. Despite the tendency to eulogise trust uncritically, there are times when others' public avowals of trust are unwelcome. Precisely because avowing trust is a way of creating a sense of obligation in others, declaring reliance on or indebtedness to others can be a ploy for establishing a relationship where none is wanted. Trust is often welcome – perhaps even usually so—but not always.

5. *Forms of trust*

Contexts in which the Ur-notion of trust is apposite, then, result in the term 'trust' being used in such a way that it acquires additional, richer resonances of meaning: of hope and of threat. Yet we are inventive language users, and happily appropriate terms for use in contexts analogically related to their original conditions. This is equally the case for trust, such that we must recognise plural, analogically related variants of the concept.

Recall that the Ur-notion of trust is generated by the need to refer to individuals *acting* in a way reliant on others' cooperative behaviour. In simple creatures, action may be a hardwired response to stimulus in the environment. But with more complex cognitive abilities, a person's action is preceded by practical deliberation and decision. Once the cognitive ability to support deliberation and decision is in place, it is a small step to describe as trust those mental states which are important constituents of the process leading to actual reliance on another's freely cooperative

action. Further, these same mental states may be present in cases where I do not actually rely on another, but where I would be disposed to do so under different circumstances, as in the rich aunt case. So the term 'trust' may be felicitously used to describe mental states that result in dispositions to rely on cooperative behaviour, as well as actual instances of reliance. The sorts of mental state that may lead people to have a disposition to rely on others includes beliefs about what will lead the other party to be trustworthy. But there is no reason to suppose that it is restricted to beliefs. Loving someone may prompt a disposition to trust, and a very robust one at that, often surviving despite evidence of untrustworthiness. And this is what we observe. The different uses of the word canvassed in §2 above show 'trust' used in ways that refer to cognitive, conative and affective mental states. All of these may be significant in issuing in a disposition to trust; so all of these are felicitously described as trust. Call beliefs which lead to a disposition to trust, *cognitive trust*. Call judgments, decisions, intentions and resolutions which lead to a disposition to trust, *conative trust*. Call emotional states which lead to a disposition to trust, *affective trust*.

Recognising the plurality of these forms of trust permits a diagnosis of where the three prominent ways of thinking about trust, also canvassed in §2 above, go wrong. Each way of thinking takes some relational situation as paradigmatic, and then builds an account of trust around that. In seeing everyone as fundamentally self-interested, the rational choice theorist is cursed with the ever-present suspicion of the imminent breakdown of relationship. For Coleman, trust situations are 'problematic' (1990, p. 92). So it is no surprise that this approach issues in an account on which trust is a belief about trustworthiness, as Hardin claims, for beliefs are pre-eminently rationally assessable.

Others are more impressed by the general stability of trust than by the imminent threat of breakdown, and so focus on those motivations that ensure well-functioning relationships. Normative accounts are especially well placed to explain more distant relationships in a communal setting, where interactions with people with whom I have no special tie with need to be governed by impersonal rules of interaction. In these contexts, the normative imperatives are an excellent basis for trust. Affective accounts take intimate relationships as paradigmatic, where love and goodwill between persons is substantially what motivates mutual support and taking each other into account. Indeed, Baier is explicit that her purpose is to displace an emphasis on relations between free and equal strangers, which she finds implicit in moral philosophy's contractarian tradition, with an emphasis on the frequently non-chosen relationships of dependence in the home (Baier, 1994, p. 114). All of these three strands of thinking about trust—the rational, affective and normative—are plausible. All make the same mistake of seizing on one form of disposition to

trust, or motivations to trustworthiness, and making these definitional of the concept.

There is a second noteworthy way in which repeated analogical use of the term extends its meaning further. Trust arises in interpersonal contexts. Because it is never certain that a person will act cooperatively, however, it is a de facto feature of trust that it always occurs in situations where, from the subjective standpoint of the putative truster, it is possible that they will be let down. Sometimes, however, we rely not on other people, but on things. While things are clearly not capable of cooperative action, it may nonetheless be opaque to us whether they will prove reliable, and their unreliability may affect us. When I lay the branch of a tree out over the ravine to bridge it, I may not know whether it will take my weight. The consequence of its unreliability is significant and costly to me. These facts allow the analogical extension of the term 'trust' to describe my walking out on the branch. In this way, I can trust things, not just people. It is an ameliorated sense of 'trust', to be sure, but it is indubitable that we do sometimes use the term in this context, with connotations of exposure to risk and uncertainty of outcome. Call this *predictive trust*, for trust here involves nothing more than a prediction of reliability, and specifically no expectation that the trusted may take account of me in their action.[5] As an after-word on this, it is noteworthy that I can adopt this stance of (merely) predictive trust towards other people, when I rely on them in situations where it would be absurd to expect them to take me into account as they act.

For clarity, I have represented the differences between trust as it arises in its generative conditions and these discrete notions of cognitive trust, affective trust, conative trust and predictive trust by labelling them distinctly. The fact that in each of these situations we would naturally apply the term 'trust' without adjectival qualification, however, illustrates my point. Because the analogies are so close, and the use of the term 'trust' so apt to describe them, repeated use has hardened these into discrete notions. So we can talk about plural forms of trust, in addition to the richer resonances acquired by trust in the central situation which gives rise to the concept, of social action. There is no reason to suppose that I have identified all the forms of trust; those that I have noted are merely the most obvious forms.

6. *Conclusion*

The genealogical account I have given, then, provides a clear explanation of why trust is so valuable to us, tending to trump other values in conversation, yet is so frequently an invisible assumption. It explains why trust eludes definition, and diagnoses why three of the dominant and diverse

strands of thinking about it have found it plausible to stress the features that they have. And it explains on how trust can come to have such varying forms. What is the significance of this genealogical account of trust? In summary, the position I have defended is quietist.

There are a number of philosophical questions that arise in relation to the concept of trust, both because of the intrinsic interest of the topic, and also because it is so fertile a perspective from which to approach different topics related to the way we live together. One enquiry, for instance, concerns the moral status of trust, whether it is permissible, or obligatory, or praiseworthy. Relatedly, another enquiry concerns the reasons which determine when trust is permitted, or obligated, or praiseworthy; Karen Jones calls this 'the justification conditions of trust' (Jones, 1996, p. 4). Another concerns whether trust is voluntary or involuntary, with the corollary of whether one could be held responsible for trusting or distrusting another. A further enquiry addresses whether there is a justified presumption of trust, or whether we require evidence before trusting rationally. A reason for elucidating trust is the hope that an analysis of the concept will answer some of these questions. Yet the principal result of this genealogical enquiry is the claim that this hope is forlorn. There is nothing about 'the concept of trust' *per se* that dictates an answer to these questions, because there is no single concept to address. The genealogical account is quietist, in the sense that it is not committed to an answer on any of these questions.

This quietist result does, however, have a more positive implication for these more substantial enquiries. Although there is no single concept of trust which, when read off the world, tells us the answer to these questions, there are some distinctive *forms* of trust which are particularly noteworthy, some of which I have identified. Stipulating that *this* is the concept which is on the table for a particular discussion provides a starting point for addressing these other questions. Once this is done, however, there is no reason to be suspicious of an analysis of that particular form of trust. So a further moral of the story is that often the simple invocation of trust will fail to be sufficiently perspicuous. Clarity in philosophical debate requires specifying which type of trust is at issue.

These points are well illustrated by Paul Faulkner's use of trust as a justification of hearers' acceptance of testimony (2007a, developed in his 2011). Faulkner is explicit that both actions and attitudes can be described as trust, and he too adopts Hollis's distinction between predictive and normative trust (2011, pp. 23–5). So he implicitly recognises that there is no monopoly on how the word 'trust' may be used. He then offers an analysis of the kind of trust as an attitude that he is interested in, which he terms 'affective trust': A trusts S to φ 'if and only if (1) *A* depends on *S* φ-ing; and (2) *A* expects (1) to motivate *S* to φ (where *A* expects this in the sense the *A* expects it *of S* that *S* be moved by the reason to φ given by (1))'

(2011, p. 146; the analysis is thus a development of the normative approach to trust initiated by Holton). It is not stated by Faulkner what the status of this analysis is—must *all* instances of attitudinal trust which are more than mere prediction conform to it, or is he concerned solely to pick out a form of trust which does exist? If the former, the analysis is subject to counter-example; that his conditions are not necessary is shown by the trust-between-mother-and-son case, where normative expectation is absent or idle.[6] So charity in interpretation suggests the latter, a reading which is hardly forced given his willingness to talk elsewhere of 'this sense of trust' (2007b, p. 311). What is indubitable is that his analysis picks out a phenomenon which is aptly described as trust; I doubt counter-examples could be produced to show his conditions not to be sufficient. And this is all that he requires. Given that his analysis does describe an actual phenomenon, Faulkner is able to do some interesting work in the epistemology of testimony, a related but nonetheless different topic to my present one. His work shows well how, with clarity on the notion of trust, philosophical progress can be made.[7]

Sidney Sussex College
Cambridge, UK

NOTES

[1] The defences of the method are in Craig, 1990, Section I and Williams, 2002, Chapter 2.

[2] Other important genealogical approaches include Jane Heal's elucidation of the notion of rationality (Heal, 2007) and Paul Faulkner's explanation of our practice of giving and believing testimony (Faulkner, 2007b, 2011). The title of his 2007b, 'A Genealogy of Trust', suggests that our approaches may be rivals. I do not think this is so. The principal subject of his genealogy is our testimonial practice, in explaining which he uses the notion of trust. (This is explicit in his later monograph, where sub-section §7.4 is entitled 'A Genealogy of Testimony'; 2011, pp. 188–97.) In contrast, the subject of my genealogy is the concept of trust. Nonetheless, Faulkner proposes that his genealogy of testimony also provides a genealogical justification of the value we place on trust (personal communication). For we must value trustworthiness intrinsically in order for our testimonial practice to be possible, and the practical value of a broadly successful practice of giving and believing testimony is such that this justifies an intrinsic valuing of trustworthiness (the thought derives from Williams, 2002, pp. 84–100). In reply: this justifies an intrinsic value being ascribed to trustworthiness; I do not see that it justifies the same intrinsic value being ascribed to trust. I comment on Faulkner's analysis of trust in §6 below.

[3] Competition can occur in a broader context of cooperation, of course. A boxer trusts his opponent not to come at him with a knife, although there is no trust when it comes to the matter of landing punches on each other.

[4] The phenomenon is observed by Dasgupta (1988, p. 53); Gambetta (1988, p. 234); Pettit (1995) and Jones (1996, p. 22), from whom I take the term.

[5] I take the label from Hollis, 1998, p. 10.

[6] Faulkner (personal communication) argues in rejoinder that my mother-and-son case is not a counter-example. For the son knows that the mother loves him and knows that this will

motivate her to care for his needs. So conditions (1) and (2) of his analysis of affective trust are fulfilled. In reply: this illustrates the ambiguity of 'to expect' in English. The verb may be used in the sense of *requiring* of someone that they φ, by holding them to an obligation. Or it may be used in the sense of *predicting* that a person will φ. If used in the analysis in the normative sense, then I maintain that the original case is a counter-example. If 'expect' is used in the predictive sense, as in his rejoinder, then a modified case serves as a counter-example. In this, suppose that the son is loyal to an abusive mother whom he has every reason to believe will not be motivated to care for him by his dependence on her. But regardless of her dismal track record, and believing that he will likely be disappointed, he hopes that she will be so motivated. His attitude is permissibly described as 'trust' yet does not meet Faulkner's conditions. So they are not necessary for all cases of non-predictive attitudinal trust.

[7] I am grateful to Alex Oliver, Jane Heal, Paul Faulkner, Tim Crane, Jeroen Van den Hoven, Hallvard Lillehammer, Richard Harper, Richard Holton, Josh Hordern, Sam Kimbriel, Harry Adamson and an anonymous referee for discussion of issues and criticism of earlier versions.

REFERENCES

Acton, H. B. (1974). *The Idea of a Spiritual Power*. London: Athlone Press.

Baier, A. C. (1994). *Moral Prejudices: Essays on Ethics*. Cambridge, MA: Harvard University Press.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: Belknap Press.

Craig, E. (1990). *Knowledge and the State of Nature*. Oxford: Clarendon Press.

Dasgupta, P. (1988). 'Trust as a Commodity', in D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, pp. 49–72.

Defoe, D. (1719). *The Life and Adventures of Robinson Crusoe*. References to 1965 edn., A. Ross, ed. Harmondsworth: Penguin.

Faulkner, P. (2007a). 'On Telling and Trusting', *Mind* 116, pp. 875–902.

Faulkner, P. (2007b). 'A Genealogy of Trust', *Episteme* 4(3), pp. 305–321.

Faulkner, P. (2011). *Knowledge on Trust*. Oxford: Oxford University Press.

Gambetta, D. (1988). 'Can We Trust Trust?' in D. Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, pp. 213–237.

Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Heal, J. (2007). 'Back to the Rough Ground! Wittgensteinian Reflections on Rationality and Reason'. *Ratio* 20(4), pp. 403–421.

Hollis, M. (1998). *Trust Within Reason*. Cambridge: Cambridge University Press.

Holton, R. (1994). 'Deciding to Trust, Coming to Believe', *Australasian Journal of Philosophy* 72, pp. 63–76.

Hume, D. (1748). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. References to 1975 edn., L. A. Selby-Bigge, ed. and P. H. Nidditch, rev. 3rd edn. Oxford: Oxford University Press.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Jackson, F. (2005). 'Ramsey Sentences and Avoiding the *Sui Generis*', in H. Lillehammer and H. Mellor (eds) *Ramsey's Legacy*. Oxford: Clarendon Press, pp. 122–136.

Jones, K. (1996). 'Trust as an Affective Attitude', *Ethics* 107, pp. 4–25.

Nickel, P. J. (2007). 'Trust and Obligation-Ascription', *Ethical Theory and Moral Practice* 10(3): 309–319.

O'Neill, O. (2002). *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press.

Pettit, P. (1995). 'The Cunning of Trust', *Philosophy and Public Affairs* 24(3), pp. 202–225.
Smith, A. (1790). *The Theory of Moral Sentiments*. References to 2002 edn., K. Haakonssen, ed. Cambridge: Cambridge University Press.
Strawson, P. F. (1974). 'Freedom and Resentment', in his *Freedom and Resentment and Other Essays*. London: Methuen, pp. 1–25.
Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.