Trust and Risk

Philip J. Nickel and Krist Vaesen

To appear in:

Handbook of Risk Theory, Roeser, S. (ed.), Springer Publishers.

Final draft: please do not quote without permission

Abstract: Philosophical conceptions of the relationship between risk and trust may be divided into three main families. The first conception, taking its cue from Hobbes, sees trust as a kind of risk assessment involving the expected behavior of another person, for the sake of achieving the likely benefits of cooperation. The second conception of trust sees it as an alternative to calculative risk assessment, in which instead of calculating the risks of relying on another person, one willingly relies on them for other reasons, e.g., habitual, social or moral reasons. The third conception sees trust as a morally-loaded attitude, in which one has a moral expectation that one takes it to be the responsibility of the trusted person to fulfill. In the context of interpersonal relationships this attributed moral responsibility creates spheres perceived to be free of interpersonal risk, in which one can pursue cooperative aims. In this chapter we examine how these three views account for two *prima facie* relationships between risk and trust, and we look at some empirical research on risk and trust that employs these different conceptions of what trust is. We then suggest some future areas of philosophical research on the relationship between trust and risk.

I. Introduction

This chapter describes some prominent philosophical connections between trust and risk. Section II briefly describes historical thought about these connections. In Section III, we describe current philosophical research on trust, including relevant empirical work on trust, attempting to understand how trust helps people pursue welfare cooperatively under conditions of risk. We conclude by indicating, in the final section, areas of future philosophical research on trust and risk.

II. History

Trust thought of as a practical solution to situations of risk has its modern origins in Hobbes.[1] Humans outside of civil society, Hobbes argues, are vulnerable to risks both from the harsh natural environment, and to risks from other people. It is rational to protect one's own life, Hobbes argued, but for this very reason rational cooperation is not generally possible, because in many situations it will be rational to exploit others' willingness to cooperate, and to expect similar exploitation from others. Therefore outside of social and political structures it is rational to treat all other persons in a warlike way, cooperating with them only if it is possible to guarantee that they will not do harm to oneself. Many cooperative goods will then be impossible to obtain (Hobbes 1968 [1651], I.13). For this reason, Hobbes thinks it is practically necessary to establish a powerful political authority that will coercively enforce the terms of promises and ensure the possibility of mutual cooperation (*ibid*., I.14). Hobbes calls *trust* the attitude of rational confidence in the reciprocal cooperation of others that is safeguarded by such a powerful authority (*ibid*., I.15).

The particular way that trust figures in Hobbes' political philosophy was not addressed explicitly by his many contemporary detractors (they focused more on his conception of the rational person and the natural law). Indeed, the notion of trust was relatively unimportant in moral, political and social philosophy throughout the modern period.[2] When trust reemerged as a philosophical topic in the Twentieth Century, it did so in three main areas. Some scholars picked up where Hobbes left off, seeking to understand how cooperative relations going against immediate self-interest but beneficial for social well-being are strategically and motivationally enabled under conditions of risk (Held 1968,

[1] Hardin (1998, 9) points out that Hobbes' conception of trust has origins in the "anonymous Iamblichi" (1995), an ancient Greek author. A somewhat similar thought is expressed by the character Protagoras in Plato's dialogue *Protagoras* (Taylor 1976, 14).

[2] An exception is its appearance in the political philosophy of John Locke, a fact which is emphasized as particularly important in some interpretations of his political views (Laslett 1988, Dunn 1984).

Coleman 1990). A second group sought a more wide-ranging understanding of trust as an explanatory concept in social theory and social science, thinking of trust as an emergent property of societies in modernity (Luhmann 1979). Finally, a third group, beginning with the work of Annette Baier, sought a moral-philosophical account of trust. Baier criticizes the other traditions for failing to give a satisfactory normative ethical account of trust outside of contractual or quasi-contractual relations. She points out that trust relationships are characteristic, perhaps especially so, of relationships that cannot easily be thought of as contractual or exchange-based, e.g., parent-child relationships and other relationships among intimates, and that here and elsewhere the question of whether trust is good or bad is not merely strategic (Baier 1986). Since all three of these conceptions of trust are still actively relevant to the relationship between trust and risk, they will be discussed in the next section on current research.

III. Current Research

A. Philosophical Conceptions of Trust and Risk

In the first subsection below we give some initial definitions of risk and trust, and then set out two apparent conceptual relationships between risk and trust. This will provide an impression of the relationships to risk that should be explained by a philosophical theory of trust. We then proceed to discuss how three families of philosophical conceptions of trust explain and interpret these relationships.

A.1. Two relationships between trust and risk

Whereas the concepts of risk and risk-perception are clearly-defined, the concept of trust is relatively problematic and unsettled. Following Hansson, we define risk as a possible but not certain future harm, or the probability of such a harm, or the expected disutility of

such a harm (Hansson 2004). Risk perception is the mental representation of a risk, as realized in emotions such as fear and in cognitive states such as prediction.[3]

Trust is a concept used variously by research in psychology, sociology, management studies, philosophy, economics and political science. There are many conflicting accounts of what it is, how it arises, and what kind of explanatory work it might perform in an account of interpersonal relations. Even within these fields there is disagreement about how to define and investigate trust (McLeod 2002, Hardin 2006). Because of these disagreements, we will settle on as neutral an initial definition of trust as possible. Trust is *a disposition willingly to rely on another person or entity to perform actions that benefit or protect oneself or one's interests in a given domain*. This definition departs from some prevalent accounts of trust in several ways. First, it leaves out the distinctively moral character that the attitude of trust is often conceived to have by philosophers. (This view will be discussed in subsection D below.) Second, it fixes two objects of trust: the agent, and the actions-within-a-domain that are to be performed by the agent.[4] In this way it is more concrete than accounts of trust that view the object of trust as a measure of general confidence in the good intentions of other persons or institutions. The third point relates to the *willingness* of trust. When one's disposition to rely on another is only due to coercion or a lack of alternatives, it is not a trusting disposition. In cases where one trusts, one has some independent reason for doing so, relating to the qualities of the trusted entity. The definition thus leaves open what characteristic reasons support trust, whereas most accounts of trust specify this in various

[3] It is sometimes useful to draw a distinction between risk and epistemic uncertainty. Risk is one kind of uncertainty: a quantified or well-characterized uncertainty. Other kinds of uncertainty are less well quantified or characterized. For example, in a poker game, the likelihood that a given poker hand will lose is a well-characterized possibility, given a fair game. The likelihood that the hand will lose because the game is not fair, however, might not be easy to characterize or quantify. We sometimes want to distinguish these two factors by calling the former a risk and the latter an epistemic uncertainty.

[4] This is referred to in the literature as "three-place trust" (Baier 1986), and is already implicit in one of the earliest contemporary philosophical treatments of trust (Horsburgh 1961). Sometimes the third element is a domain of activity, and at other times it is a particular (type of) action.

ways, for example by saying that it is characteristically based on social information, such as group-membership, relationship, reputation, and power.

This last point is somewhat subtle, but important for our understanding of the relationship of trust and risk. It is the first of two main conceptual relations we will set out in order to see how theories of trust address the issue of risk. Suppose we are thinking of hiring WDC Dredging Company to conduct a small dredging project. The *Handbook of Dredging Engineering* states that "Characterizing the environmental risk posed by dredged material requires information on the likelihood that organisms will be exposed to contaminants and the probability that such exposure will produce adverse effects in the environmental receptors (organisms) of concern. The uncertainty associated with these estimated risks has been high" (McNair 2000, 22.39). Special techniques have been developed to reduce these risks under different conditions. In order rationally to rely on WDC to carry out the project, however, we need not quantify these risks and uncertainties ourselves, reaching a conclusion about how they should be addressed. We could instead confidently count on WDC to know and apply appropriate methods, and to abide by relevant laws and industry practices — in a word, we could trust them. Our reasons for this would have to do with WDC's reputation, with the existence of regulatory bodies, and with the knowledge that WDC wishes to protect its future reputation. They would not normally depend on specific information about probable dangers due to this particular dredging job.

This suggests that the question whether it is a good or bad idea to rely on another person or entity to perform a given action can be answered in two ways. One way is by evaluating the risks associated with that action. The other is by determining whether the other person or entity is trustworthy. For example, whether it makes sense to rely on WDC to dredge a harbor safely and without environmental damage partly depends on the quantifiable risks that are associated with this agent performing this activity. But it also depends on the

trustworthiness of WDC relative to other companies. Therefore, in a situation like this, risk evaluation and determination of trust are in some sense *different answers to the same question*: whether to rely on the company to perform a given action. Whereas risk evaluation focuses on the probabilities associated with the underlying activity, determination of trust focuses on the qualities of the entity (in this case, the company) such as its competence and motives, as influenced by its social and normative standing. In case one of the answers is incomplete or inadequate, it may be possible to answer the question in the other way: where risks cannot be easily quantified under the circumstances, one can ask whether the company is trustworthy; where the company is unknown, one can attempt to quantify the risks independently. This is echoed in a recent survey article by a group of psychologists: "There are two general ways [that individuals predict cooperation in others]. One, *trust*, is defined as the willingness, in expectation of beneficial outcomes, to make oneself vulnerable to another based on a judgement of similarity of intentions or values; the other, *confidence*, is defined as the belief, based on experience or evidence, that certain future events will occur as expected" (Earle *et al.* 2007, 30).[5]

A corollary is that actively controlling risks tends to reduce trust rather than increasing it. To the extent that we attempt to quantify and control the risks and uncertainties of this dredging job and choose the proper methods ourselves, this reduces the scope of our trust in WDC by reducing the extent of our reliance and vulnerability, even though the action that we essentially rely on them to do — dredging the harbor — remains the same. If we attempt to reduce our risk by taking out insurance, this reduces the extent to which our well-being depends on WDC's fully adequate performance. Or suppose we want to be as sure as possible that WDC will perform safely, and to this end we do a detailed investigation into this company's entire history of activity and the safety of the methods they use, and moreover

[5] Luhmann (1979) and Seligman (1997) also draw a distinction between trust and confidence. Given Earle *et al*'s definition, it might make sense to define trust as a very special kind of confidence, as Deutsch (1977) does.

formulate a plan to monitor the stages of their dredging operation. This would reduce our epistemic reliance on WDC. In such cases where we take active steps to control or reduce our risks, epistemically and practically, it might be appropriate to remark that we *do not fully trust* WDC, or at any rate that after going through these steps we do not need to trust them as much. It seems, then, that as our own knowledge and control of the risks associated with reliance on another entity increase, our independence increases, our reliance and vulnerability decrease, and our (need for) trust in the other entity decreases. Hence in this way, too, the characteristic reasons for risk-taking and for trusting are different from one another. This, then, is the first main conceptual point about risk to be explained by theories of trust: the characteristic reasons for risk-evaluation and for trust are different, and to some extent mutually exclusive.

The second main apparent conceptual relationship to be explained by a theory of trust is that high background risk makes trust more salient, but also more difficult to justify. Suppose Q is the organizer of a protest group under a repressive government. Q faces the risk that if the group is discovered, she will go to prison. In order to expand membership in the group and coordinate group actions, she must tell others about the group and count on them to carry out activities on the group's behalf in strict confidence. In such a situation Q must rely on other people to protect her interests and those of the group. Her willingness to do this goes along with her trust. It seems that in cases like this where risks are high, but where reliance on others is essential to a valued activity, well-grounded trust is of great importance (Boon and Holmes 1991, cited in Das and Teng 2004, 87). Trust often coexists with risk and is made more salient by it, because the need for trust increases as the risks associated with a valuable act of cooperation increase. On the other hand, in trusting one wants to minimize risk as much as possible by finding the most competent, cooperative people available and the most conducive background conditions. In our example, Q would be

rational to trust those individuals least likely to compromise her interests while helping fulfill her goals and those of the group. It might be rational, for example, to collaborate with family members or close associates of people already in the group, because these persons have more to lose if they defect. Because the stakes are high for Q, she must be very careful about whom she trusts and under what circumstances. Thus, the risks associated with a cooperative act make trust more important, but also more difficult to adopt on careful, stringent grounds.

One way of justifying this point is to draw on intuitions from recent epistemology about the relation between knowledge-ascription and practical stakes. The intuition is that the greater the practical stakes that hang on the truth of some claim P, the more difficult it is to have knowledge or sufficient justification for believing P. Here is a version of a standard example used to motivate this intuition:

> It's late on Friday afternoon, and raining hard. Lo and her next-door neighbor Hi are [separately] thinking about going out to the bank, but wondering whether the trip could be postponed until tomorrow. Both of them can remember a recent Saturday visit to the bank, but neither of them has any further information relevant to the question of whether the bank will be open tomorrow. Nothing much is at stake for Lo —her banking errand could be done any time in the next week—but for Hi the question has burning practical importance. Hi knows he must deposit his paycheck before Monday, or he will default on his mortgage and lose his home. (Nagel 2008, 279; the original example is found in DeRose 1992)

Among epistemologists, a widely-adduced intuition about this case is that whereas Lo knows (or is justified in believing) that the bank is open on Saturday, Hi does not know this, even though Hi and Lo have the same evidence. The only difference between Hi and Lo is what they have at stake concerning the question whether the bank is open, hence this is taken as the key difference that explains why one of them has sufficient justification for knowledge and the other not. High-stakes practical reasons raise the bar of sufficient justification. Assuming we can extrapolate from this type of high-stakes case to other cases, such as Q's question whether a given person will keep secrets about the protest group from the authorities, it

appears that sufficient justification for trust is difficult to come by partly because the epistemic standards for knowledge or justified belief are higher in Q's high-stakes situation.[6]

However, there is also the more mundane principle that where more is at stake, there is often more reason for any given individual relied upon not to perform as expected, and consequently more reason not to trust them. People in Q's country take a much greater personal risk by participating in a protest group than those in countries without an oppressive government, so they need to have special motives or qualities such as courage in order to perform as Q expects of them. What is needed for trustworthiness in this high-stakes context may therefore be more demanding. Since the characteristics underlying trustworthiness are more rare, it will be more difficult for Q to justify trust.[7]

These, then, are two main apparent conceptual relationships between risk and trust that one hopes a theory of trust will elucidate: on the one hand, extensive risk evaluation makes trust less relevant; on the other hand, when the risks are greater, trust is more difficult to justify. In the remainder of this section we will set out three prevalent approaches to the theory of trust and see how well they explain and justify these apparent relationships.

A.2. Trust as interpersonal staking

The first approach under consideration sees trust as a special instance of rational risk-taking. As Sztompka writes, "Trust is a bet about the future contingent actions of others" (1999, 25; see also Gambetta 1988). Suppose Q considers inviting Ron to join her protest

[6] Some have doubted whether this intuition is shared by many people, and have conducted survey studies that disconfirm it (Buckwalter 2010). Nagel (*op cit*.) grants the intuition, but argues that it is explained not by high practical stakes, but by "need-for-closure," the psychological need that people feel to settle a question, for example when there is time pressure or distraction. Notably, these recent discussions take the claim as an empirical psychological truth, not a normative epistemological principle. In this sense it cannot directly settle normative theoretical questions about when we should trust, nor can it justify people's trust, but it can only explain why people trust the way they do.

[7] It is also distinctive of this example that those who are trustworthy to Q may have special reason to conceal the very characteristics that would indicate this to others. There also may be government agents who pretend to have such characteristics in order to infiltrate and undermine the protest group. Thus Q's epistemic task is difficult indeed.

group. If Ron is trustworthy, his membership in the group will bring a significant added benefit to Q. If Ron is not trustworthy, inviting him could be very detrimental to Q. Thus, in considering whether to invite Ron, Q must weigh the likely benefits of inviting him against the possible risk that if she does so, Ron will betray her and the group. This can be regarded as a calculated choice, or a kind of bet Q is in a position to make on Ron's future behavior. Some have regarded the willingness to make such an interpersonal bet as the essence of trust. There are important methodological reasons for considering trust in this way, such as its ability to help predict and explain macro-level social effects as a function of the cooperative behavior of individuals with limited options and interests, where the values of different options for individuals (including the option of reliance on others) can be thought of as preference rankings or subjective expected utilities (Coleman 1990, Ch. 1).

This type of view does an excellent job explaining the second of the conceptual points set out above, why the difficulty of justifying trust increases with the risks. This is due to the obvious fact that on such a view one's dispositions to trust are based on expected utilities, which take into account both the likelihood of performance, and the stakes to be gained through performance and lost through non-performance. The increasing benefits of performance can increase the willing disposition to rely on another; similarly, the increasing harms of non-performance can decrease this willing disposition. This means that the willing disposition is not exclusively dependent on the *belief that another person will perform* (is likely to perform). It is just as dependent on the *stakes* that are attached to performance and non-performance, where the expected utility of reliance given these stakes is compared with a situation in which one relies on somebody else, or does not rely on anybody. Such a view is therefore quite different from a doxastic (belief-based) view of trust according to which the willing disposition is based on a *belief* that the entity relied upon will perform in a certain way. One need only believe it is likely enough that the entity will perform to make it

worthwhile, given the utilities of performance and non-performance, to rely on him (Nickel 2009). In cases where there is little to lose given non-performance, one need not believe strongly that a person will perform in order to feel comfortable relying on him. But that is not Q's situation. She has a lot to lose through Ron's non-performance, so before she can willingly rely on Ron she needs information that more fully ensures that Ron will perform, and this increases the justificatory burden for a belief in his trustworthiness.

However, this view does not do much to explain the first of the conceptual points set out above, the mutual exclusivity between reasons for trusting and detailed risk assessment. To emphasize this, consider a thought-experiment involving the dredging job again. Suppose I have conducted detailed research on the history of all dredging jobs in my region, including those of WDC. I also come to learn almost perfectly the scientific appraisal of the methods used by WDC. I come to have as much knowledge about these topics as anybody ever has before. Suppose, too, that I have taken out a large insurance policy to protect me in case WDC does something wrong for which I am liable. I also form a detailed plan for monitoring WDC. Now on the "staking" view of trust that we've been considering, this gives me perfect grounds for the attitude of trust: the more I can guarantee performance and safeguard against non-performance, the more purely I can trust. But this is exactly the opposite of the first intuitive link discussed above, according to which these are not the characteristic reasons for trust.

Some advocates of this type of approach have specified that trust relates only to reliance undertaken on the basis of certain sorts of reasons. One prominent view in particular adopts the condition that Q's trust is characteristically based on Q's belief that the one to be relied upon (Ron, in this case) cares about Q's interests as part of Ron's own interests: Q believes that her interests are *encapsulated* in Ron's (Hardin 2006). This moves some way toward explaining the distinctively interpersonal nature of reasons for trust. According to this

encapsulated interest view of trust, possible reasons for thinking one's own interests are encapsulated in another person's interests include the thought that it is for their own strategic future benefit to perform as expected, that they are subject to sanctions if they do not perform as expected, or that they are fair and law-abiding. The risks involved in trusting are interpersonal risks: they therefore depend on, among other things, the motives of the trusted person.

However, Hardin does not hold that efforts to quantify risks objectively or control them result in a reduction of trust. Indeed, on his view they would seem to enhance trust, since trust is for him still calculative. Reacting against this view, the sociologist Eric Uslaner distinguishes between strategic trust and moralistic trust (2002). Strategic trust is based on rational assessment and verification; in moralistic trust, one relies on others without detailed risk assessment and without expecting anything else in return, as in Uslaner's example of a fruit stand owner who leaves the stand unoccupied, expecting others (mostly people he does not know and will never meet) to pay on the honor system for what they take (*ibid.*, 14–15). There are many situations in which we seem to trust unknown others by default, without frequent disappointment (Verbeek 2002, 140–1). Decisions to trust have been shown not to be based on risk information in some studies (Eckel and Wilson 2004). Indeed, some have even argued that we are under some kind of weak moral obligation to regard others in this way, until they are proven unreliable (see Thomas 1978). The question then is what, if anything, could make it epistemically rational to adopt a trust attitude on these grounds. So as not to lose a connection with a quantitative behavioral theory, Das and Teng suggest that this appearance of non-calculative, default trust is an illusion: estimates of the likelihood of a trusted person's performance can be posited below the surface as "unconscious calculations" about the goodwill and competence of others, based on estimates of risk (2004, 99).

A.3. Trust beyond rational calculation

The second approach sees trust as an interpersonal belief that is essentially non-rational or non-calculative, involving to some extent a leap of faith. Such a view can originally be traced to Luhmann (1979), who held that trust is primarily a means of reducing complexity by relying on others instead of making an independent assessment of risk. As Guido Möllering sets out the idea, many philosophers of a rational choice orientation are led a certain distance in this direction by the need to acknowledge the occasional therapeutic or future-oriented, rather than evidential, nature of trust. Placing trust is sometimes used as a means of cultivating reciprocal engagement and cooperation; in many cases it can actually help produce the effect of performance, even though there was no specific prior evidence this would occur. In this way trust is "reflexive" (Möllering 2006, 80–84). But according to Möllering, this does not go far enough, because this still assimilates trust to a calculation based on past performance and future consequences, now taken in a slightly broader sense. Möllering advocates a more radical account that instead views trust as more fundamentally non-rational, according to which it is truly a "leap of faith" in which one suspends calculations about benefits and risks to some degree. On this view, trust does not just reduplicate an explanation that could already be provided by the rational choice paradigm, but can instead be usefully applied to empirical phenomena of cooperation that do not easily fit that paradigm, such as patients' suspension of rational considerations when electing a surgery (*ibid.*, 188–9). Another expression of this idea is that trust involves *optimism* about the performance of others (Jones 1996), an attitude in which despite the presence of risk a good outcome will be achieved (Nooteboom 2002).

If we are reluctant to postulate such a motivation for cooperation beyond a subjectively rational basis, then we can explicate such a view more prosaically by claiming

that the essential reasons of trust are substantially interpersonal, social and/or moral instead of calculative.[8] Thus trust is properly based on the following kinds of considerations:

- the desire of the trusted to engage in future interaction with the trustor;
- the desire of the trusted to protect their future reputation (Pettit 1995);
- the security of the institutional context in which the trusted person operates;
- the moral qualities of the trusted person (see subsection 4 below);
- general confidence that things will work out no matter what happens;
- the ethical desirability of trust itself.

That some of these are reasons for trust is already acknowledged by advocates of the staking model. But their acknowledgement is subject to the understanding that such elements can be converted into calculative factors in one's judgment of the benefit and likelihood of different outcomes. Taking the reasons in this way disregards one of the principle motivations of the view under discussion, which is that trust is "a functional *alternative to rational prediction* for the reduction of complexity. Indeed, trust is to live as if certain rationally possible futures will not occur" (Lewis and Weigert 1985, 976, italics added; cited in Möllering 2006, 83). Certain risks are therefore excluded from consideration at the moment of deciding to trust. On such a conception perfectly central, normal cases of trust would include situations in which no good calculative basis for trust is available, nor is employed (in terms of a concrete track record or an estimation of general reliability in similar situations), but willing reliance on another nonetheless occurs.

Such a view does a good job explaining the first of our earlier intuitive claims about the relationship between trust and risk, that trust-formation and risk-evaluation seem like quite different processes, and that risk-evaluation and control tend to make trust irrelevant or out of place, rather than increasing it. Since trust is non-calculative, risk-evaluation and control tend to push it aside. As for the second of our earlier claims about risk and trust, the non-calculative view of trust takes a paradoxical stance. On the one hand, it grants the

[8] It is important not to conflate this non-calculative view of trust with the view that trust is emotive. On the one hand some emotions, e.g., fear, can be regarded as affective 'calculations' of risk, and on the other hand some socially-based judgments have none of the affect characteristic of emotional states.

normative claim that from a restricted rational choice point of view, justifying trust becomes more difficult as the stakes increase. But it holds that trust can proceed without this type of justification, either on the basis of sheer optimism, or on the basis of interpersonal, social, or moral reasons. In order to have an interesting field of application in the real world, such a conception of trust invites the corresponding empirical hypothesis that the occurrence of willing reliance on others is not directly inversely proportional to the magnitude of risks perceived to be associated with this reliance. Evidence on this point is ambivalent. Empirical studies of the prisoner's dilemma and other similar game situations seem to indicate that people sometimes willingly rely on others even in the absence of calculative reasons to do so (Möllering 2006, 38; Hardin 2006, 50–1). This suggests that such a non-calculative conception of trust may well have an interesting domain of application.

However, there are still philosophical worries about such a view. For example, Pamela Hieronymi has argued that the attitude of trust is conceptually constrained by the fact that it inherently answers the question "Is the other person trustworthy?" (Hieronymi 2008). In much the same way as belief in P answers the question, "Is it the case that P?" and an intention to φ answers the question, "Is φ to be done?", the attitude of trust toward some person R answers the question, "Is R trustworthy?" Although Hieronymi leaves open what kinds of considerations close the question of whether R is trustworthy, her account implies that in really adopting the attitude of trust, one is always committed to having sufficient reason for thinking that the other person is trustworthy. To the extent that her argument is plausible, it tethers the notion of trust to the reasons for thinking another person trustworthy. Genuine trust is not compatible with the presence of severe, realistic risks associated with reliance that are known to the truster and have not been ruled out, controlled or hedged.

A4. Trust as a moral attitude

The third approach to the concept of trust sees it primarily as a moral attitude among intimates and community members, defining social spheres within which cooperative actions are safely pursued. A good example of this approach can be seen in the work of Caroline McLeod, who thinks that any notion of trust should first accommodate "prototypes" or central examples such as "child-parent relationships, intimate … relationships, and professional-client relationships" (2002, 16), and then address theoretically useful but more peripheral examples such as self-trust and trust between strangers. According to McLeod, the prototypical examples of trust are best seen as involving an ascription of certain moral and relationship-regarding qualities to the trusted person, such as moral integrity, and a shared perception of the relationship itself. These qualities ensure that the person will have *goodwill* toward us, a characteristic belief of trust first emphasized by Annette Baier (1986). The focus on trust as connected with the quality of intimate relationships can be traced back to the developmental psychologist Erik Erikson (1950), who held that the quality of the mother-child bond determines trust-dispositions that are crucial for normal social development.

Baier holds that trust essentially involves remaining vulnerable to the actions of another person, partly due to the discretion given to that person within a domain of activity (Baier 1986). This vulnerability is especially present in intimate relationships for several reasons: because the domain of shared activity is large; because the discretion given is considerable; and because the relationship itself is often valued intrinsically, so that something extra is at stake should a failure or breakdown occur. This is sometimes manifested in a disposition to *feelings* of vulnerability as well, something noted by empirical studies of partner trust (Kramer and Carnevale 2001). On Baier's view, we willingly increase our vulnerability to another because we are confident of the *competence* and the *goodwill* of the other person, although we often don't consider these factors consciously. Vulnerability

provides a link with risk on Baier's account. Correspondingly, feelings of vulnerability or security provide a link with risk-perception.[9]

The reason philosophers often claim that the attitude of trust has an essential moral dimension is that trust seems intuitively to commit one to certain characteristic morally-tinged responses in cases where another person fails to perform as they were trusted to do. Moral philosophers are careful to distinguish two fundamentally different kinds of expectation that we can take toward other people when we rely on them (Holton 2004, Faulkner 2007). *Predictive* expectations are simple beliefs that the future will go one way or another. They do not inherently involve a personal or emotional commitment to its going one way or another. When one relies on the future being a certain way in this predictive sense, one is at most disappointed if one's expectation is not met. *Normative* expectations, on the other hand, involve a prescriptive judgment or assumption that the future *should* go a particular way. If the future doesn't go this way — if a person doesn't do what she has agreed to do, for example — then moral disapprobation and other richly evaluative or moral judgments are appropriate. In cases where one willingly relies on another person, but is not disposed toward any moral judgment such as blame or betrayal toward them if they do not perform, intuitively this does not seem like a central case of trust. It is *mere reliance*, a simple prediction that a person will behave a certain way, not trust.

In its refined form, then, trust seems to have a moral dimension. Philosophers have differed in how they describe this moral component, however. Some philosophers have characterized it as a requirement that there should be mutually shared values between the trusting person and the trusted person (McLeod 2002). Others have suggested a weaker

[9] Vulnerability (as a state, rather than a feeling) is conceptually related to risk. In talking about the causal components of risk, it is common in technical literature on risk management to distinguish between vulnerability, exposure, and hazard (Renn 2008, 69). Vulnerability is a (sometimes relational) property of a thing that makes it potentially susceptible to damage or harm, should exposure to a hazard source occur. Baier does not draw this distinction, but it is instructive to do so. We might restate Baier's view by saying that trust involves, not increasing one's inherent vulnerability, but instead increasing one's *exposure* to a potential hazard source — another person — by giving them discretion over something important.

condition than that of shared moral values, since it seems that I can trust somebody even when I know them to have rather different values and standards than myself. It has been variously suggested that in trusting I commit myself to blaming a trusted person if he or she does not perform (Holton 1994), holding them responsible for performing (Walker 2006), or taking a moral expectation toward them (Nickel 2007). None of these suggestions requires that the trusted person share the values of the trusting person, only that they be sufficiently responsive to moral requirements and situational expectations that they are suited to be the objects of moral attitudes.

If we think of trust as a morally loaded attitude, what does this mean for the relationship between trust and risk? On this view, trust presupposes an awareness of moral expectations and responsibilities in the trusted person, making that person more trustworthy and reducing the human risks in interpersonal reliance. If Q, from our earlier example, has as part of her trust a moral expectation of Ron, the content of which is that he keep information about the group confidential, this presupposes a view of Ron on which he is capable of responding to the morally salient features of the situation, including the vulnerable situation of Q and her group. Thus, if Q is right in her moral expectations, this minimizes the purely interpersonal, strategic risk in relying on Ron, allowing them jointly to focus their attention on their other goals within their cooperative activity, including the avoidance of other risks.

On such a view, one would often expect to see extensive risk evaluation of a person during the process in which trust is established, but once it has been established this activity would largely cease, allowing one to turn one's attention elsewhere. (However, in some relationships, e.g., parent-child relationships, trust will simply be presupposed rather than established.) Continuing surveillance is therefore a sign that trust has not yet been achieved or that it has failed (Baier 1986). This partially captures the first intuitive point discussed earlier in section III. A.1, according to which extensive scrutiny of risks is incompatible with

trust. However, it also leaves room for relationships in which full trust has yet to be established or is never reached, so that risk evaluation is still ongoing.

Our second intuition above was that when the risks are greater, trust is more difficult to justify. On moral conceptions of trust, the attitude of trust is partly justified by one's reasons for thinking that the trusted person has certain moral qualities, or is capable of responding to moral expectations. We could therefore look to this feature to explain why trust is more difficult to justify under conditions of risk. Social psychology has shown that people's tendency to do the virtuous or morally required thing *toward a person unknown to them* is highly susceptible to background conditions such as time pressure (Darley and Batson 1973). This suggests that under adverse conditions, moral responsibility is not usually sufficient to generate altruistic or cooperative behavior. Highly risky situations may also create circumstances in which people's moral characteristics are less reliable toward unfamiliar persons. Thus trust in strangers would be difficult to justify on the basis of attributions of moral responsibility. On the other hand, trust toward intimates or people within one's community might be much better justified under such conditions. The moral dispositions of close associates might be expected to "kick in" especially under conditions of risk and adversity, making them even more trustworthy under those conditions. Thus, here too, the moral view of trust only partly supports our earlier intuitive association between risk and trust. (It is important to note here that there are striking instances in which trust in strangers under highly adverse conditions also turns out to be justified, e.g., those discussed in Monroe 1996.)

B. Scientific approaches to trust and risk

The three conceptual approaches to trust discerned above (trust as interpersonal staking, trust as non-rational, and trust as moral attitude) are also useful for characterizing and subdividing the scientific discussion concerning the development and evolution of trust.

In what follows, we reconsider these three approaches, looking at the scientific research carried out within each paradigm. The consistency and explanatory interest of this scientific research may help to settle which approach is correct, or it may imply that we should take a pluralistic view of the nature of trust for empirical purposes.

B.1 The science of trust as rational risk-taking

Trust as risk-taking is at the heart of many game-theoretic approaches to trust: one tries to determine the factors inhibiting and enabling the evolution of trust, given *calculative self-interested agents*. For instance, many scholars have used prisoner's dilemmas as a way of studying trust. In the classic prisoner's dilemma,[10] it is in the interest of the individual to forego cooperation, and to defect. The aim, then, is to identify the conditions under which the payoffs of cooperation start to outweigh the payoffs of individual action. If cooperation does emerge, the implication from cooperation to trust seems straightforward: one can expect an agent to cooperate only when she trusts the other to reciprocate her cooperative act.

Morton Deutsch was one of the first to study these ideas empirically and systematically (Deutsch 1960). He looked at how individual trust orientations (cooperative, individualistic or competitive) affected the likelihood of individuals cooperating, and found that individuals with a cooperative general trust orientation were also more likely to make cooperative choices in prisoners dilemma situations. [11]

The main problem of Deutsch's approach (and of his followers), is that it doesn't say much about the conditions under which trust develops, unless one takes the cooperative

[10] In the classic prisoner's dilemma, two players can choose between two moves, either "cooperate" or "defect". Each player gains when both cooperate, but if only one of them cooperates, the other one, the defector, will gain *more*. If both defect, both lose, but not as much as when an agent's cooperative act is defected by the other. Since there is a risk of greater loss in case of cooperation (namely, when the other player defects), the best in terms of safety, but suboptimal, strategy is to defect.

[11] Deutsch used standard two-player nonzero-sum games, in which players were requested to imagine they were oriented either cooperatively (before playing, players were asked to consider themselves to be partners), individualistically (players were asked to play just to make as much money as possible for themselves) or competitively (players were asked to play just to make as much money as possible for themselves and also to do better than the other). Deutsch found that cooperation was most likely to develop when both players' trust orientations were cooperative.

behavior of a person as an indication of that person's disposition to trust, as Deutsch did. Under that condition, however, the exercise becomes circular, since the measure of trust (cooperative behavior) is exactly the feature that trust was said to cause (Cook and Cooper 2003). In other words, when an agent makes a choice in a prisoner's dilemma, she does two things: she decides whether or not to trust the other, and simultaneously, whether to honor the possible trust placed in her. Consequently, her decision to cooperate doesn't need to reflect her trust, for it might just as well indicate her willingness to reciprocate the other's possible act of trust.

For these reasons, scholars started to look for new game-theoretic protocols in which trust and reciprocity could be dissociated. One such protocol, currently widely used in economics and psychology, is the so-called trust-honor game (see e.g. Snijders and Keren 1999). Basically it is a sequential prisoner's dilemma, with players no longer making decisions simultaneously, but consecutively. First, player 1 has to decide whether or not to trust player 2, without knowing what the latter will decide; second, player 2, with knowledge of the first player's decision, decides whether to honor the trust by reciprocating.

To get an idea of how trust-honor games work, consider the so-called investment game, developed by Berg et al (1995). Two players get $10 each. The game consists of two stages: the trust *placing*, and the trust *honoring* stage. In the first stage, player 1 needs to decide how much money to pass to an anonymous player 2. All the money which is passed is increased by a factor greater than 1, say, 3. So if player 1 passes $10, player 2 will receive 30$, resulting in the allocation ($0, $40). In the second stage, player 2 needs to decide whether or not to honor the trust placed in her: she gets the opportunity either to keep all the money, or to send an amount x back to player 1. For all x's smaller than 30 but greater than 10, both players will be better off than in the original allocation ($10, $10). For an increase in

payoff for either player to occur at all, however, the first stage of the game is crucial: player 1 needs to be willing to a take a risk on player 2.

In the study of Berg et al (1995), players were not only anonymous, their interaction was one-shot (taking away the possible effects of reputation-building), and there was no opportunity for them to punish dishonored trust. Notwithstanding these conditions, the researchers found that their subjects did both place trust (by sending money) and honor trust (by sending money back). From this they conclude that "reciprocity exists as a basic element of human behavior and that this is accounted for in the trust extended to an anonymous counterpart" (122). Reciprocity is basic because it cannot be directly reduced to or explained in terms of calculative self-interest. Reciprocity with an anonymous partner is interesting because one cannot use similarity or other identifying characteristics to gauge future performance. One's reason for reciprocation needs to depend on the mere idea of the game situation or of an agent in that situation.

Follow-up studies changed the incentive structure of the investment game, added mechanisms of reputation-building and punishment, made the encounters between players non-anonymous, looked at cultural background effects, and so on, but one premise remained constant: trust is measured by looking at the *risk* a player is willing to take on the other. In the set-up above, player 1's behavior would indicate a high level of trust in case she passed $10 to player 2, a low level in case she didn't pass any money. Thus, Cook and Cooper (2003) write, "First, I decide whether or not to "trust" you (or take a risk on you) and cooperate on the first play of the game, then you must decide whether or not to "honor" that trust by cooperating in turn. Clearly, this behavior can be viewed as risk taking by the first player" (p. 217).

So in these studies, an explanation of trust simply *is* an explanation of the conditions under which subjects decide in favor of social risk-taking. Social-risk taking, in turn, is

calculative, as argued in Section III.A.2: one takes a risk (i.e. one trusts) in the hope of benefiting afterwards (receiving more than the $10 one started with). If trust is a strategy of generating higher payoffs for individuals, and if these individuals are calculative utility maximizers, trust games would offer a reasonable explanation indeed for why in such essentially self-interested agents as humans, an (apparent) prosocial trait such as trust (or cooperation, more generally) could have evolved.

It is worth briefly discussing the selective pressure commonly invoked here — something which will prove useful for understanding Section B.3. The prehistorical background against which the evolution of cooperation is usually set is a world that was becoming increasingly harsh. In particular, early humans started to cooperate to cope with (i) increased climatic and seasonal variability (Potts 1998; Ash and Gallup 2007); (ii) the colonization of new environments, requiring new, cooperative modes of foraging (Kaplan et al 2000); and/or (iii) increased intraspecies competition, given the expansion of hominin populations (Humphrey 1976; Alexander 1989; Dunbar 1998). In these challenging environments, an individual's self-interest was best served by the individual's contributing toward the interest of the group. From a selective point of view, cooperation succeeded where individual efforts began to fail.

The prime cooperative interaction to be explained in these evolutionary scenarios is between genetically *unrelated* agents, viz. how unrelated agents began to form bonds to settle down and defend new environments, or how some 1.5 million years ago hunter *Q* started to trust hunter *R* not to comprise *Q*'s life by dropping out of the hunting coalition early, and to do his fair share of stone-throwing and clubbing. In other words, it is assumed that genetically *related* individuals have a natural incentive to cooperate, and that therefore, non-kin cooperation has explanatory prevalence over cooperation among kin. Although this methodological choice has yielded neat models and explanations of why trust and other

prosocial behaviors *might* have developed in humans, it has obscured some of the non-calculative aspects of trust, which are discussed in the next two subsections.

In summary, in models that assume utility maximizing agents, it is common to treat trust simply as the willingness of an agent to take a risk on another agent. The aim, then, is to establish the conditions under which these calculative agents may develop such calculative form of trust. The prime explanandum of these models is economic cooperation between unrelated agents, while neglecting cooperation among kin.

B.2 The science of trust as non-calculative

There are two strands of empirical research which question the strong link between trust and risk-taking as pictured in staking accounts. The first relates to a remark made in our conceptual discussion of trust as interpersonal staking (Section III.A.2), namely that, contrary to what staking accounts would predict, extensive risk reduction (e.g., through extensive risk assessment) should decrease rather than increase levels of trust. Empirical evidence, now, indeed suggests that if one takes away the risk of defection, one takes away the substrate for trust.

In empirical research, risk reduction is usually introduced by letting agents make formal binding agreements about the transaction in which they are about to engage (see e.g., Molm et al 2000, 2009). These are often called *negotiated* exchanges: one negotiates the conditions of the transaction, and secures them through some binding agreement (as is the case in most contemporary market exchanges). Both partners know in advance what each will give and get; the agreement ensures that both partners live up to their promise. These exchanges contrast with so-called *reciprocal* exchanges. The latter are riskier in that partners, at the moment of performing an altruistic act, do not know whether the act will be reciprocated in the future. The altruistic act, one hopes, will elicit altruism in the other, although there is no guarantee to that effect.

Trust, Molm and colleagues found, is more likely to develop in reciprocal than negotiated exchanges.[12] In negotiated exchange, one might be *confident* that the other will do as promised, yet remain skeptical about the other's trusthworthiness. In reciprocal exchange, by constrast, trustworthiness is tested through real interaction; one gets the opportunity to infer the other's trustworthiness from her behavior rather than from the promises she makes. While interacting, partners typically come to trust each other more, evaluate them more positively, and feel more committed to the relationship. To be sure, both negotiated and reciprocal exchange may yield a fair reallocation of goods. Nonetheless, they have different extra payoffs: increased levels of certainty versus increased levels of trust.

Often the contrast is framed somewhat differently, namely by drawing on the distinction between cognition-based and affect-based trust (see McAllister 1995). On this account, both negotiated and reciprocal exchanges may yield trust, but of different kinds. Negotiated exchange involves cognition-based trust, which is based on beliefs and estimations of a partner's reliability, and which is abandoned in case the partner behaves differently from how she was expected to behave. Conversely, reciprocal exchange results in affect-based trust. Here beliefs about the other's reliability are complemented with affective states of care and personal regard, and often in addition with a sense of forgiveness for occasional instances in which the partner dishonors the trust placed in her.

The second line of empirical research that undermines the idea of trust being merely calculative suggests that risk-taking and trust-placing rely on two different psychological mechanisms. Note what we should expect if they didn't: risk-aversive people should exhibit a lower propensity to engage in relations of trust, and conversely, risk-minded people should be more trust-prone (Ben-Ner and Putterman 2001). Several studies have tried to establish this link, but mostly without success (see e.g., Eckel and Wilson 2004; Ashraf et al 2007; Fehr

[12] Molm and colleagues didn't use a behavioral measure of trust (e.g. cooperative behavior), but rather an attitudinal one. That is, they asked subjects about their general commitment and trust toward their exchange partner.

2009; Ben-Ner and Halldorsson 2010). For instance, in the most comprehensive study, Ben-Ner & Halldorsson (2010) looked for correlations between several measures of trust and several measures of risk-orientation. Measures of trust included both behavioral measures (i.e. how subjects behaved in an investment game) and so-called survey measures (determined by means of a questionnaire), indicating a subject's general sense of trust in other people. Also with respect to risk attitudes, both behavioral and survey measures were used. Ben-Ner & Halldorsson found that no measure of risk attitude correlated with any of the measures of trust used.[13]

That risk-taking and trust-building depend on different psychological mechanisms is also suggested by recent studies on the psychological and behavioral effects of certain hormones. Testosterone, for example, has been shown to increase risk-taking (Apicella et al 2008), but to decrease trust (Bos et al 2010). Oxytocin has precisely the opposite effect: it decreases risk-taking, but increases trust (Kosfeld et al 2005).

In sum, both strands of empirical research point to a shortcoming in attempts to characterize trust in terms of risk-taking: staking accounts largely ignore the profoundly interpersonal and social aspects of trust. There might be good methodological reasons to do so (e.g., to keep models tractable), but it is important to bear in mind that one might miss out on some salient characteristics of trust. Yet the research described here shares with staking accounts the concern of explaining trust primarily among strangers (or non-kin). That is what sets them apart from the approach considered next.

B.3 The science of trust as a moral attitude

As we remarked at the end of Subsection B.1, it is common to argue that human hypersociality evolved to cope with environments that became ever more challenging. On

[13] In contrast, gender and personality (extroversion, openness, agreeableness, conscientiousness, neuroticism) did predict trusting behavior.

this account, trust is basically a strategy to safeguard one's own fitness indirectly. By enabling cooperation among non-kin, trust increases the fitness of the collective.

Recently, this dominant view has come under attack. Sarah Hrdy (2009), for example, argues that human cooperative behavior did not evolve in a context of competition (with other groups of humans), but rather in a context of cooperative childcare among kin and as-if kin. The idea is that the delayed maturity characteristic of humans requires childcare that cannot be provided by mothers alone. A large fraction of the 13 million calories that are needed to rear a modern human from birth to maturity (approx. age 15–16), for instance, is provided not by the mother, but by other caregivers. In contrast to the great apes, human fathers and older siblings invest fairly extensively in their offspring and younger siblings. Even more crucial is the role of grandmothers: they take over when the mother is too busy, when the mother dies early, or when the mother is simply giving birth to the next infant. According to Hawkes, O'Connell, Blurton Jones, and colleagues, this explains the unusual trait of humans to live actively many years after the birth of their final child (Hawkes et al 1999). Menopause is an adaptation enabling longer childhoods, and in virtue thereof, increased braininess.

Cooperative childcare (or cooperative breeding, as it is often called) is dependent on three factors. First, the child must be able to elicit attention. It must display its vulnerability, and try to solicit the care not only of its mother, but also of more remote caregivers. Second, the mother must be able to delegate some of her work. She must be willing to put her baby's fate in the hands of others. In other words, she must trust her helpers, and assume their moral qualities, viz. good intentions and selfless motives. In fact, human mothers are unique in this respect. Unlike the other great apes, human mothers do not hold jealously onto their infants, but willingly allow others to hold them. A human mother has a default sense of trust: she assumes—in the absence of counter-evidence—that others will not harm her helpless child. If

these strangers did harm the baby nonetheless, there would be a clear sense of wrong-doing, backed up by moral or proto-moral attitudes. Third, both the mother and her helpers (grandparents, fathers, siblings, and other as-if kin) must sense and be responsive to the vulnerability and to the particular needs of the infant. Mothers are particularly well-adapted on this score, as they tend to be attuned to the slightest perturbations in the conditions of their young, and to perceive all neonates as attractive (Hrdy 2009). Arguably, such a positive attitude towards babies applies to the human species in general, since it is also commonly found (in varying degrees) in male adults, teens, adolescents, and so forth.

In the view of authors such as Hrdy, the prosocial behavior exhibited in cooperative breeding forms the cornerstone of our morality. In the context of cooperative breeding, early humans developed their capacity to form relations of trust with kin and as-if kin, a capacity which was extended to non-kin only later. In conclusion, from a scientific perspective it seems to make good sense to view trust as an attitude that originates in our dealings with intimates and community members—as expressed by Carolyn McLeod in Section 4.A. It implies a sphere in which even the most valued things, such as one's child, can be safely entrusted to others. Trust in cooperative breeding is the prototypical form of trust, from which other more peripheral examples of trust derive.

IV. Areas of Future Research

In this final section we set out a few areas for future philosophical research. As suggested in the previous section, there is still empirical research to be done within the conceptual paradigms of trust. "Empirical philosophy" could contribute to this project. Recent philosophical "experiments" have provided us with additional information about people's intuitive conceptions of intentionality (e.g., Knobe 2003), weakness of will (Mele *forthcoming*), causation (Livengood and Machery 2007), and consciousness (Knobe and

Prinz 2008), so there is no principled reason why similar methods could not yield novel insights regarding (folk) conceptions of trust. Such research may eventually suggest that one of the conceptions of trust is more theoretically interesting or useful than the others, or that there is a way of reconciling them. Such work should take up the relationship of trust and risk as an important challenge.

Second, greater philosophical development of some of the underlying conceptual paradigms would be useful to the understanding of human responses to risk. The non-calculative conception of trust stands in need of greater philosophical conceptualization, a point that Möllering also makes (2006, 126). Much of the conceptual work on this theory of trust has been carried out by sociologists. Philosophers may also have something to contribute to such an account of trust. It may also be useful to ask whether the moral account can be reconciled with either the calculative or the non-calculative account of trust, explaining how moral considerations make trust more strategically rational, or how they help people to bypass the need for strategic rationality.

Finally, an area of particular interest in future philosophical work is understandinghow trust is related to technological risk. For several reasons, trust in technology has become a key issue in recent years. First of all, as Meijboom (2008, 31) points out, food production and other means of satisfying basic welfare have become increasingly large-scale, decentralized and socially and technologically complex. In this context, perception of technological risk is prevalent as a background fear. Ulrich Beck (1992) famously argues that the imposition of technological risks has reached a critical juncture, at which we must reconsider our practices of justifying and imposing such risks. Horror and science fiction films in which technology gives rise to bad consequences are evidence that this fear resonates with the public. Risk is increasingly impersonal. As Renn puts it, “personal experience of risk has been increasingly replaced by information about

risks, and individual control over risk by institutional risk management. As a consequence, people rely more than ever on the credibility and sincerity of those from whom they receive information about risk" (2008, 222). However, Onora O'Neill suggests that the solution to fears of technological risks does not lie in making more data available to the public. Paradoxically, she points out, greater openness has not created greater trust in the UK health care system, for example (2002a, 72–3; 2002b, 134–5). After all, the public has a difficult time processing information about risks, and does not have time to assess all the data with respect to every technology they could rely upon. These remarks suggest that trust is crucial to understanding public attitudes toward technological risks. It could be that a philosophical approach emphasizing the centrality of trust to the use of technology would help reconcile technocratic and public attitudes about technological risks.

In this chapter, we have shown that although the relationship of trust and risk is still unsettled, and is grounded in underlying disagreement about the nature of trust, that disagreements about this relationship can be given a clear characterization. This may lead to such disagreements eventually being settled by further methodological reflection and empirical study, or it may just mean that different paradigms of trust can be used for different purposes in thinking about risk. In that case, a sketch of the available paradigms for thinking about this relationship, such as we have given here, will be of some value.

References

"Anonymus Iamblichi" (1995) In Gagarin M, Woodruff P (eds) Early Greek political thought from Homer to the Sophists. Cambridge University Press, Cambridge, pp 290–295

Alexander R (1989) Evolution of the human psyche. In: Mellars P and Stringer C (eds) The human revolution: Behavioural and biological perspectives on the origins of modern humans. Princeton University Press, Princeton NJ, pp 455–513

Apicella C, Dreber A, Campbell B, Gray P, Hoffman M, Little A (2008) Testosterone and financial risk preferences. Evolution and Human Behavior 29:384–390

Ash G, Gallup G (2007) Paleoclimatic variation and brain expansion during human evolution. Human Nature 18:109–124

Ashraf N, Bohnet I, Piankov N (2007) Decomposing trust and trustworthiness. Experimental Economics 9:193–208

Baier A (1986) Trust and antitrust. Ethics 96:231–60

Beck U (1992) Risk society: Towards a new modernity Trans. Ritter M. Sage Publications, Thousand Oaks, CA

Ben-Ner A, Halldorsson, F (2010) Trusting and trustworthiness: what are they, how to measure them, and what affects them. Journal of Economic Psychology 31:64–79

Ben-Ner A, Putterman L (2001) Trusting and trustworthiness. Boston University Law Review 81:522–551

Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity and social history. Games and economic behavior 10:122–142

Boon SD, Holmes JG (1991) The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In: Hinde RA, Groebel J (eds) Cooperation and prosocial behavior Cambridge University Press, Cambridge: 190–211

Bos P, Terburg D, van Honk J (2010) Testosterone decreases trust in socially naïve humans. Proceedings of the National Academy of Sciences 107:9991–9995

Buckwalter W (2010) Knowledge isn't closed on saturday: a study in ordinary language. Review of Philosophy and Psychology 1: 395–406

Coleman JS (1990) Foundations of social theory. Harvard University Press, Cambridge, MA

Cook K and Cooper R (2003) Experimental studies of cooperation, trust and social exchange. In: Ostrom E, Walker J (eds) Trust and reciprocity: Interdisciplinary lessons from experimental research. Russell Sage Foundation, New York, pp 209–244

Darley JM and Batson CD (1973) 'From Jerusalem to Jericho': A study of situational and dispositional variables in helping behavior. Journal of Personality and Scoial Psychology 27:100–108

Das TK and Teng B (2004), The risk-based view of trust: A conceptual framework. Journal of Business and Psychology 19:85–116

DeRose K (1992) Contextualism and Knowledge Attributions. Philosophy and Phenomenological Research 52:913–29

Deutsch M (1960) The effect of motivational orientation upon trust and suspicion. Human Relations 13:123–139

Deutsch M (1977) The resolution of conflict: constructive and destructive processes. Yale, New Haven

Dunbar R (1998) The social brain hypothesis. Evolutionary Anthropology 6:178–190

Dunn J (1984) The concept of trust in the politics of John Locke. In: Rorty R , Schneewind JB, Skinner Q (eds) Philosophy in history: Essays on the historiography of philosophy. Cambridge University Press, Cambridge, pp 279–302

Earle TC, Siegrist M, Gutscher H (2007) Trust, risk perception and the TCC model of cooperation. In: Siegrist M, Earle TC, Gutscher H, Trust in cooperative risk management: Uncertainty and scepticism in the public mind. Earthscan, London, pp 1–49

Eckel CC, Wilson RK (2004) Is trust a risky decision? Journal of Economic Behavior and Organization 55:447–465

Erikson EH (1950), Childhood and Society. Norton, New York

Faulkner P (2007) On telling and trusting. Mind 116:875–902

Fehr E (2009) The economics of trust. Journal of the European Economic Association 7:235–266

Gambetta D (1988) Can we trust trust? In: Gambetta D (ed) Trust: Making and breaking cooperative relations. Basil Blackwell, London, pp 213–237

Hansson SO (2004) Philosophical perspectives on risk. Techné 8: 10–35

Hardin R (2006) Trust. Polity, New York

Hardin R (1998) Trust in government. In: Braithwaite VA, Levi M (eds) Trust and governance. Russell Sage Foundation, New York, pp 9–27

Hawkes K, O'Connell J, Blurton Jones N, Alvarez H, Charnov E (1999) Grandmothering, menopause and the evolution of human life histories. Proceedings of the National Academy of the Sciences 95:1336–1339

Held V (1968) On the meaning of trust. Ethics 78: 156–9

Hieronymi P (2008) The reasons of trust. Australasian Journal of Philosophy 86:213–236.

Hobbes T (1968) Leviathan [orig. 1651], Macpherson CB (ed) Penguin, New York

Horsburgh HJN (1961) Trust and social objectives. Ethics 72:28–40

Hrdy S (2009) Mothers and others: The evolutionary origins of mutual understanding. Harvard University Press, Cambridge, MA

Humphrey N (1976) The social function of intellect. In: Bateson P, Hinde R (eds) Growing points in ethology. Cambridge University Press, Cambridge

Kaplan H, Hill K, Lancaster J, Hurtado AM (2000) A theory of human life history evolution: Diet, intelligence, and longevity. Evolutionary Anthropology 9:156–185

Knobe J (2003) Intentional action and side effects in ordinary language. Analysis 63:190–194

Knobe J, Prinz J (2008) Intuitions about consciousness: Experimental studies. Phenomenology and the Cognitive Sciences 7:67–83

Kosfeld M, Heinrichs M, Zak P, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. Nature 435:673–676

Kramer RM, Carnevale PJ (2001) Trust and intergroup negotiation. In: Brown R, Gaertner S (eds) Blackwell handbook of social psychology: intergroup processes (Blackwell, London, pp 431–50

Laslett P (1988) John Locke, two treatises of government. Cambridge University Press, Cambridge

Lewis JD, Weigert A (1985) Trust as a social reality. Social Forces 63:967–985

Livengood L, Machery E (2007) The folk probably don't think what you think they think: Experiments on causation by absence. Midwest Studies in Philosophy 31:107–127

Luhmann N (1979) Trust and power: Two works by Niklas Luhmann. Wiley, New York

McAllister D (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal 38:24–59

McLeod C (2002) Self-trust and reproductive autonomy. MIT Press, Cambridge, MA

McNair EC (2000) The U.S. Army Corps of Engineers dredging operations and environmental research (DOER) program. In: Herbich JB (ed) Handbook of dredging engineering. McGraw-Hill Professional, New York, pp 22.01–22.46

Mele A (2010) Weakness of will and akrasia. Philosophical Studies 150: 391–404

Meijboom FLB (2008) Problems of trust: A question of trustworthiness. Dissertation, Utrecht University

Möllering G (2006) Trust: Reason, routine, reflexivity. Elsevier, Amsterdam

Molm L, Schaefer D, Collett J (2009) Fragile and resilient trust: Risk and uncertainty in negotiated and reciprocal exchange. Sociological Theory 27:1–32

Molm L, Takahashi N, Peterson G (2000) Risk and trust in social exchange: An experimental test of a classical proposition. American Journal of Sociology 105:1396–1427

Monroe KR (1996) The heart of altruism: Perceptions of a common humanity. Princeton, Princeton, NJ

Nagel J (2008) Knowledge ascriptions and the psychological consequences of changing stakes. Australasian Journal of Philosophy 86: 279–294

Nickel PJ (2007) Trust and obligation-ascription. Ethical Theory and Moral Practice 10:309–319

Nickel PJ (2009) Trust, staking and expectations. Journal for the Theory of Social Behaviour 39:345–62

Nooteboom B (2002) Trust: Forms, foundations functions, failures and figures. Edward Elgar Publishing, Cheltenham, UK

O'Neill O (2002a) A question of trust: The BBC Reith lectures 2002. Cambridge University Press, Cambridge

O'Neill O (2002b) Autonomy and trust in bioethics. Cambridge University Press, Cambridge

Pettit P (1995) The cunning of trust. Philosophy and Public Affairs 24:202–225

Potts R (1998) Variability selection in hominid evolution. Evolutionary Anthropology 7:81–96

Renn O (2008) Risk governance: Coping with uncertainty in a complex world. Earthscan, London

Seligman AB (1997) The problem of trust. Princeton University Press, Princeton, NJ

Snijders C, Keren G (1999) Determinants of trust. In: Budesco D, Erev I (eds) Games and human behavior: Essays in honor of Amnon Rapaport. Lawrence Erlbaum, Mahwah, NJ

Sztompka P (1999) Trust: A sociological theory. Cambridge University Press, Cambridge, NJ

Taylor CCW (1976) Plato Protagoras. Oxford Clarendon Press, London

Thomas DO (1978) The duty to trust. Proceedings of the Aristotelian Society 79:89–101

Verbeek B (2002) Instrumental rationality and moral philosophy: An essay on the virtues of cooperation. Kluwer Academic Publishers, Dordrecht

Walker MU (2006) Moral repair: Reconstructing moral relations after wrongdoing. Cambridge University Press, Cambridge