Trust, Staking, and Expectations

Philip J. Nickel

[Draft Version. Final version at *Journal for the Theory of Social Behaviour* 39 (2009): 345–362.]

Abstract: Trust is a kind of risky reliance on another person. Social scientists have offered two basic accounts of trust: *predictive expectation* accounts and *staking* (betting) accounts. Predictive expectation accounts identify trust with a judgment that performance is likely. Staking accounts identify trust with a judgment that reliance on the person's performance is worthwhile. I argue (1) that these two views of trust are different, (2) that the staking account is preferable to the predictive expectation account on grounds of intuitive adequacy and coherence with plausible explanations of action; and (3) that there are counterexamples to both accounts. I then set forward an additional necessary condition on trust (added to the staking view), according to which trust implies a moral expectation. When *A* trusts *B* to do *x*, *A* ascribes to *B* an obligation to do *x*, and holds *B* to this obligation. This Moral Expectation view throws new light on some of the consequences of misplaced trust. I use the example of defensive medicine by physicians to illustrate this final point.

Keywords: Interpersonal trust, moral expectations, prediction, rationality, defensive medicine

I. Introduction

Social scientists are interested in trust because levels of trust vary a great deal among different societies and groups and are thus an interesting object of empirical study, and because trust is a kind of "social capital," a social resource that makes cooperation and mutual benefit easier to obtain (Coleman, 1990; Field, 2003: 62–5; Fukuyama, 1995; Herreros, 2004). In this paper I explain and criticize two prevalent social scientific accounts of trust, which I call Predictive Expectation and Staking Accounts. I then propose an alternative account on which trust is a moral attitude. I argue that only on such an account can we capture the importance of blame and betrayal to trust.[1] My alternative view clarifies and strengthens the claim that trust is a powerful means of achieving cooperation and social compliance.

The target concept here is sometimes called "three-place" trust (Baier, 1986: 236) because it is a relation between persons *A* and *B* and some performance *x*. (I will throughout the paper adopt the term "performance" to refer to the successful carrying out of a relied-upon behavior.) I trust my neighbor to pick up my newspaper while I am out of town, or I trust my physician to refer me to a specialist, or I trust a hotel employee not to steal my bag. I will assume that if one person trusts another *tout court*, then the first trusts the second across the whole range of salient contexts in which they can expect to interact. But my focus will primarily be on the particular things one person trusts another to do. There is reason to prefer the three-place formulation for my purposes rather than other ways of talking about trust that de-emphasize the performance of the trusted person. When trust is broken, we often focus on the specific act by which the person violated trust. This suggests that trust has specific acts as its object, even though we may not be able to say exactly which acts we are counting on the other person to perform before the situation has evolved. In addition, the three-place formulation is assumed by the social-scientific theories I criticize.

The following partial characterization of trust is accepted by all the accounts I will discuss: trust implies *the voluntariness of risky reliance on another person (or persons).* According to this partial characterization, trust implies, first of all, a voluntary state. The trusting person is at least minimally aware of reliance and accepts it. Voluntary reliance is thus neither totally unconscious or accidental. It is also voluntary in the sense that it cannot be performed out of terror or coercion. To do so would be incompatible with trust. Second, trust is risky. The trusting person has something at stake in relying on the other person, and what is at stake cannot as a matter of fact be safeguarded with total confidence. Finally, trust is voluntariness of reliance

on another *person* (or persons). It is an interpersonal state of mind, not just any attitude of reliance toward an object that one counts on. We could consider a more general psychological state of trust synonymous with reliance, such that one could be said to trust natural objects, artifacts and people, but what is under discussion in social scientific views is instead a social concept. These features, then, are common to all the accounts of trust discussed in this paper

In this paper, for clarity of focus, I will be concerned only with cases in which both the trustee and the trusted are single, independent agents. The discussion could be extended to cases in which the trustee and/or the trusted entity is a group of two or more persons, or an institution. It could also be extended to cases in which the trusted entity is an artifact (e.g., an airplane or a software program). But such cases raise special issues of distributed responsibility, social ontology, and the ontology of artifacts, and require conceptual complexity that would distract from the main argument here. (On the question of the ontology of groups, see, e.g., Gilbert, 1989, Searle, 1995; on the ontology of artifacts, see, e.g., Meijers, 2000.)

There are three fundamental notions we could use to characterize the mental state that underlies voluntary trusting reliance on another person: predictive expectation, willingness to bet or stake something, and moral expectation. The first of these, predictive expectation, is a matter of regarding performance as more likely than not, that is, at least .5 likely.[2] People do not normally say that they predict something unless they think it is at least probable to this degree; otherwise they would say that they *suspect* it might happen, or that they believe there is a *chance* it will happen, not that they predict it. This notion of predictive expectation appears to be at work when it is said, for example, that trust is "the expectation of technically competent role performance" (Barber, 1983: 14). Niklas Luhmann writes that trust is "the generalized

expectation that the other will handle his freedom, his disturbing potential for diverse action, in keeping with his personality — or, rather, in keeping with the personality which he has presented and made socially visible" (1979: 39). More recent definitions of trust also suggest the predictive account (e.g., Lewicki et al., 1998; Newton, 2001).[3]

The second kind of attitude that might underlie willingness to rely on another person is a judgment that it is worth betting or staking something on her performance (Coleman, 1990: 99), what I will call a staking judgment (despite the artificiality of this term). According to the Staking Account, trust is based on a staking judgment that the likelihood of nonperformance times the magnitude of the associated loss is outweighed by the likelihood of performance times the magnitude of the associated gain. When talking about this judgment of willingness to bet on performance, I will speak of judging or thinking it "worthwhile" or "likely enough" to stake one's own goods on the other person's performance. This attitude is plainly different from a predictive expectation, although the difference has scarcely been noticed in the literature on trust. We must bear in mind that we are only thinking of voluntary staking, not cases in which one is coerced or forced into staking something on another person's behavior. We also must not overintellectualize or hypostasize the staking judgment, as if it were always highly calculative and deliberative.

Finally we come to the third kind of attitude, that of *moral expectation*, and the associated view on which trust implies a moral expectation. As I conceive the idea of moral expectation, it has two components: one judges it obligatory that another person perform, and one holds the other person responsible for performing. We may initially think of moral expectation in the context of a promise or contract, but in fact the notion extends more widely than this. It is a very

general type of interpersonal attitude, held by friends, intimates, family members, and strangers toward one another. Unlike promises and contracts, moral expectations are often completely tacit. One person can morally expect another person to do something simply because she has done it before, or because it is part of her expected role. At the same time, however, these moral expectations can easily be made explicit: if one person explicitly addresses another second-personally by saying, for example, "I expect you to go to the bank this afternoon," the word "expect" is typically being used to imply a moral expectation. In what follows, I will defend a view on which the central attitude of trust is a combination of the staking notion and the moral expectation notion. I will begin by considering the Predictive Expectation and Staking Accounts in more detail.

II. Predictive Expectation Accounts

Predictive expectation is a matter of regarding an event as more likely than not. So, according to a Predictive Expectation account of trust, one trusts when and only when one is willing to undertake risky reliance on another person on the basis of a judgment that performance is more likely than not. There are many variations of this view, but one prominent example can be found Russell Hardin's account of trust. According to Hardin, one person trusts another if the first has reason to believe it will be in the second's interest to be trustworthy in relevant contexts of interaction. Trust depends on a judgment of trustworthiness. In turn, this judgment of trustworthiness depends on a judgment that the trusting person's interests are "*encapsulated in the interests of the Trusted*" (2006: 19). Hardin's view is a Predictive Expectation view because of the way he goes on to explain this idea: he describes his view as "a rational expectations

account in which the expectations depend on the reasons for believing that the trusted person will fulfill the trust" (2006: 31). Hardin qualifies this by saying that "there is some risk that my interests will trump yours and that I will therefore not fulfill your trust in me; and your trust will be limited to the degree to which you think my encapsulation of your interests gives them *enough weight to trump other interests I have*" (2006: 19, italics added).[4] It is this qualification that concerns me. If the trusting person, insofar as he trusts, thinks his own interests will *trump* the other interests of the trusted, then he must regard it as (at least) more likely than not that the trusted person will perform appropriately. For example, my trust that you will use my ATM card properly while I loan it to you implies that I think you care enough for my relevant (financial) interests to trump your interest in your other projects and desires, which might otherwise lead you to take my money. This grounds a judgment that your performance is, at a minimum, more likely than not. Hence this is a Predictive Expectation view of trust.[5]

The objection to Predictive Expectation accounts of trust is that a person can take the attitude of trust toward somebody else's performance without predicting, or believing it is more likely than not that they will perform. Take two examples:

> **MARC** believes that personal physicians ought to refer patients to specialists whenever there is an uncertain diagnosis for serious medical symptoms. His experience in the past suggests that physicians hardly ever meet this obligation. But he likes his new physician and has developed a good rapport with her. Presenting to her with a complaint of severe dizzy spells, he trusts her to make a referral, even though he is reluctant to assign a likelihood to her doing so, given his past experience. If he were forced to assign a likelihood, he would say it was about a one-in-four chance.
>
> **JANET** would like to save money on childcare and often has errands that would be easier to run without her young children along. Her oldest child is fourteen, and although in the past he has not shown the maturity of judgment to be a caretaker, Janet is now disposed to put him in this position. She would trust him to take care of his younger siblings for a short period, even though she is reluctant to assign any likelihood to the claim that he will show good judgment if she does leave him as babysitter. If she were forced to assign a

likelihood, she would say it is as probable as not that he will make some significant mistake.

Marc and Janet appear to trust a given person to do some particular thing even though each is disinclined to assign a positive likelihood to that person's doing that thing, and in each case there is no track record obviously supporting a positive likelihood of performance. Therefore, it appears that Marc and Janet do not have a predictive expectation of performance. Each is disposed toward an act of reliance with some associated risk. Hence, these two cases suggest that it is possible to trust without having a predictive expectation of performance.

Despite the fact that Marc and Janet resist any explicit attempt to pin down a predictive expectation, this does not prevent them from trusting. For trust, it does not appear to matter whether a person has a belief to the effect that performance is more likely than not. What is important is whether the trusting person regards it as "worthwhile" or "likely enough" to stake his or her behavior on the behavior of the other person. Looking ahead, what I wish to emphasize for the purpose of contrasting the Predictive Expectation view with the Staking view of trust, is that *predicting* performance, and regarding it as worthwhile to *stake something on performance*, are two distinct states. On a simple approximate psychological theory of action, voluntary reliance on another person results from the judgement that the likely consequences of non-reliance are worse than the likely consequences of reliance; *not* on the judgment that performance is more likely than non-performance. The difficulty of finding a better doctor may make it more worthwhile for Marc to rely on the doctor he has even though he does not predict that she will perform as he wishes or thinks she ought. The convenience of having childcare and the disadvantages of not having it may bring Janet to rely on her teenage son even though she

does not predict he will perform as she hopes. Thus there is often no need to ascribe any very precise prediction about whether a trusted person is likely to perform. Yet trust is still possible.

Let us examine this idea in a bit more detail. Sometimes it is said that the aim of a theory of trust is to explain why people are willing to enter "covenants" in Hobbes's sense: contracts or agreements in which one of the parties "deliver[s] the Thing contracted for on his part, and leave[s] the other [party] to perform his part at some determinate time after" (Hobbes, 1968: 193; see also Hardin, 2006: 22). For example, if Al helps Bobbie string her tennis racquet, with the understanding that Bobbie will later give him tennis lessons, this is a covenant in Hobbes' sense, because Al is relying on Bobbie to deliver her "contracted" service to him at a time later than he delivers his service to her. There is an open question whether Bobbie, once she has already obtained the racquet-stringing she wants, will choose to spend her time giving tennis lessons to Al. Performance appears not to be in her own interest, since she gets nothing from it after he has already delivered his part of the bargain. Hence there is a question whether it can be rational for Al to agree to this exchange in the first place. If this is what trust is supposed to explain, then trust marks the condition in which Al finds it likely enough that Bobbie will perform that it is rational for him to enter this covenant.

With this background, the criticism of the Predictive Expectation account can be framed in a slightly different way. Suppose (again) that trust marks the condition in which A finds it likely enough that B will deliver *x* such that it is rational for A to enter a covenant with B for the sake of *x*. On this supposition, it need not be the case that A finds it more likely than not that (i.e., predictively expects that) B will deliver *x*. He need only find it likely enough to make it worth the expenditure of his resources. There is no principled lower limit to the likelihood that

makes it worth his while to enter the covenant. For example, it might be worth it to Al to string Bobbie's tennis racquet even if he estimates only a 30% likelihood of her delivering tennis lessons in return. Thus it can be worth it to him to enter a covenant with Bobbie even if he does not predictively expect that she will perform. The relevant account of trust, given the aim of explaining how covenants are possible, is a Staking Account of trust, not a Predictive Expectation Account.[6]

III. Staking Accounts

On a Staking Account of trust, the attitude of trust is a judgment that it is worth staking something of value on the voluntary action of another person. Because it is a judgment about what it is worthwhile to stake or to bet, the rationality of betting applies. Thus it is rational to trust another person when the risk of that person's nonperformance is worth taking (where the *risk* is the *likelihood* times the *cost*), given the possible benefits of performance. If, with Coleman, we define p as the probability that the person will perform in a particular way, L as the loss associated with nonperformance, and G as the gain associated with performance (as against the status quo), it is rational to trust just in case $p/(1 - p) > L/G$ (Coleman, 1990: 99). This rule serves as a norm for those who trust, a norm that is sometimes violated but not willingly and explicitly. For the sake of illustration, let us suppose that dollar units are identical to units of utility, ignoring diminishing marginal utility. Suppose Bobbie's performance is 40% probable, and the loss to Al from her non-performance is worth $100 to him, whereas the gain from her performance is worth $200 to him. In that case $p/(1 - p) = 2/3$ and $L/G = 1/2$. Since $2/3 > 1/2$, it is rational for Al to trust Bobbie. On the Staking Account, this means nothing more than that he

judges it worthwhile to stake his own actions on Bobbie's performance. His voluntary risky reliance on Bobbie is implied by this judgment that relying on her is worthwhile.

With regard to the rationality of betting, Coleman's initial statement of the norm is artificially simplified. A more complex and accurate statement of it would take into account other options or bets available to the subject. If I have a higher probability of a better outcome and a lower probability of a worse outcome by relying on person *C* rather than *B*, then it does not make sense to rely on *B* even if doing so would be a good bet by itself. Accounting for a comparative judgment of other bets (besides trusting *B* and not trusting *B*) against a common baseline, would require a more intricate statement of the norm. For simplicity's sake, Coleman's statement assumes that we are evaluating gain and loss against the next best available alternative, which we might naturally assume to be the consequences of non-trust in the status quo.

Keynes writes in his *Treatise on Probability*, "of two hypotheses it may be rational to act on the less probable if it leads to the greater good" (1921: 307). This is a consequence of an empirically plausible theory of action according to which people generally act for the sake of what they regard as the balance of costs and benefits, given what they regard as the respective likelihoods of those costs and benefits. We can think of this theory of action in a very broad sense, so that it need not be committed to the idea that all the considerations one thinks of as benefits must arise from or reduce to self-interest, nor to the idea that the benefits and drawbacks of an action cannot include moral considerations and especially considerations of justice in their own right (Sen, 1982; Smith, 2000).[7] On the Staking view of trust, trust is an action like any other, in the sense that it is normally based on an evaluation of the benefits and their likelihoods versus the drawbacks and their likelihoods. It fits with Keynes' principle. On the Staking view

the only thing that sets trust apart, as a type of action, is that it involves staking something significant on the actions of *other people*, rather than on impersonal events and consequences. Hence trust is an essentially interpersonal type of choice.

By contrast, as we have seen, the Predictive Expectation view holds that when I trust somebody to do *x* I must believe it (at least) more likely than not that she will do *x*. This starkly conflicts with the cost/benefit theory of action. Believing that something is *more likely than not* to have a desired consequence is not an essential part of explaining why a person acts a certain way, according to the cost/benefit theory. For example, in order to explain why I purchase a particular book of poetry, according to the cost/benefit theory, it is not necessary that I think it *more likely than not* that I will enjoy the book in some way. It need only be likely enough to make the cost of the book worth my money, relative to other ways I could spend that money. "Likely enough" could mean less than an even chance. The Staking account is thus preferable to the Predictive Expectation account on the ground that it coheres better with this standard theory.

Despite the merits of the Staking account of trust, however, there is a serious problem with it, which is that it is too broad and inclusive in what it regards as instances of the attitude of trust. A staking judgment, in cases of voluntary risky reliance on another person, is not sufficient for trust. Consider another pair of cases. Suppose Robert, in difficult circumstances, leaves his thirteen-year-old son with his ex-girlfriend Linda (who is not the child's mother) and moves away without announcing his new address or communicating with Linda. Suppose Robert has no standing child-care arrangement with Linda.[8] Now consider two different attitudes that Robert might have toward leaving his child with Linda:

(1a) Robert believes it is worth it to him to leave the child with Linda. If Linda fails to care for his son minimally (barring some reasonable excuse), he would blame her for it. **(Reliance + Blame)**

(1b) As in (1a), Robert believes it is worth it to him to leave the child with Linda. But in this variant, if Linda fails to care for the child minimally (barring some reasonable excuse), Robert would not blame her for it — and would not see the appropriateness of blame. **(Reliance No Blame)**

In both (1a) and (1b), Robert is disposed to stake his behavior (and his child's well-being) on Linda's acting a certain way, voluntarily engaging in risky reliance on her. Hence, according to the Staking Account of trust, both (1a) and (1b) are cases of trust. In attributing the attitude of trust, it makes no difference to the Staking Account what attitude Robert would have toward Linda in the event that she does not perform. It is concerned only with Robert's judgment about the possible payoffs and liabilities of counting on Linda.

Intuitively, however, there is an important difference between (1a) and (1b). It seems to matter, in our saying that Robert trusts Linda, whether he is disposed to blame her in the event that she does not perform appropriately (and has no excuse). (Or more accurately — though for the sake of simplicity I will not generally put the point in this way — it matters whether Robert judges it *appropriate* to blame Linda — whether or not he actually blames her.) For imagine the Robert of (1b) talking about his act of reliance on Linda: "I don't hold Linda to taking care of my son. I wouldn't blame her if she didn't, and I don't think it would be appropriate to blame her. I just think it's worth it to me to bet that she will, so that I can leave in these difficult circumstances." Now suppose that Linda decides not to look after Robert's son, and while Robert is away the child becomes a ward of the state. When Robert returns and discovers this situation, he feels no blame towards Linda, despite what has happened. Moreover, this isn't just

the result of Robert's pessimism or lack of engagement — he feels bad, even upset, about his son's not having had a good parent around, but he also judges that blame is inappropriate toward Linda. Indeed, the question of Linda's needing an excuse for her behavior doesn't even arise in his mind, since his reliance on her was just a bet. In that case, I think, it would make no sense for Robert to tell Linda that he trusted her to take care of his son. Intuitively, it is not trust but rather "mere reliance," albeit for something Robert regards as very important. Thus it appears that, conceptually, some disposition to blame Linda is necessary in order to trust her. But neither of the accounts we have examined draws any such distinction in marking out the attitude of trust.

Some may be inclined to object that there is at least a thin sense in which Robert trusts Linda even in case (1b). Consider another example. Suppose I count on my neighbor to walk past my window at the same time every day, because it reminds me to pick up my child from school. Suppose my neighbor and I have never discussed the matter and the neighbor has no independent reason to walk by my house at that time. In such a case it would be crazy to blame my neighbor if he does not walk by, even if his not doing so leads to bad consequences and makes me upset. It would also be inappropriate to tell him that I trusted him to walk by at that time, because that would almost certainly lead him to think that I blamed him for not doing so. So normally I would not blame him or tell him I trusted him. But someone might nonetheless want to claim that there is a sense in which I do trust him, where this just means that I have a kind of impersonal faith or confidence that he will do some particular thing. At any rate, people sometimes speak this way.

Such a sense of trust, however, does not play a distinctive role in the interpretation or explanation of social phenomena. The way I rely on my neighbor in such a case is no different

from how I might rely on a natural event like the afternoon rain. There is no harm in using the word "trust" to describe one's reliance in such cases, but it seems only half-serious, as in the utterance "I trusted the rain to fall this afternoon." We could interpret such a statement as fully serious by interpreting it as "I trusted *that* the rain would fall this afternoon," but in that case trust has become a bland propositional attitude indeed, a mere substitute for "believed"or "thought." In the context of social explanation, it seems best to reserve the term "mere reliance"for such a thin attitude, and use the word "trust" for cases in which the trusting person has a richer attitude.

IV. A Moral Expectation Account

In response to these considerations, I propose that trust implies a moral expectation as a necessary condition: *A* trusts *B* to do *x* only if *A morally expects B* to do *x*. This necessary condition is added to the partial set of necessary conditions set out at the beginning of the paper, according to which trust requires that *A* accepts risky reliance on *B*'s doing *x*. It is also assumed that trust involves a staking-judgment, according to which *A* regards her reliance as a worthwhile bet on *B*'s behavior (although this judgment can of course be wrong, it is still a commitment of the person who trusts). I label this set of conditions on trust the Moral Expectation account. In the first instance, it is a psychological account of the attitude of trust and is not ontologically committed to the existence of any particular moral obligations or entities. This is because it is an analysis of the attitude of trust in terms of the ascription of a moral obligation, not in terms of the actual existence of a moral obligation. But at the same time it is a moral account of trust, because of the moral quality of what is ascribed by the person who trusts. These features allow us to use the notion of trust as an explanatory tool, by helping us to explain action in social

contexts where moral norms — sometimes plural and even conflicting — are ascribed to a particular person by various parties who have something at stake. In the next section I will give an example of how this might work.

Other philosophers have proposed accounts of trust with a strong connection to moral attitudes, for example, on which trust has a strong connection to holding others responsible and having "reactive attitudes" such as a disposition to blame (Holton, 1994; the terminology is from Strawson, 1962), or on which one person trusts another only when (s)he thinks the trusted person shares his or her own core moral values (McLeod, 2000), or even on which trust always creates valid moral obligations (Hertzberg, 1988). But there has been little attempt within philosophy to argue against prominent social scientific conceptions of trust on behalf of a moral view. One of my goals in the above sections has been to provide an argument of this kind. (It goes beyond the scope of this paper to examine the rich sociological discussion of trust, morality and social explanation from Durkheim onward.)

My account of moral expectations draws on three fundamental philosophical points, each of which is put forward in Wallace's (1994) seminal discussion of them. The first and most important point is that moral expectations imply "a practical *requirement* or *prohibition* in a particular situation of action" (Wallace, 1994: 22). Here is how I understand this idea: we should think of *A*'s moral expectation that *B* will do *x* as a condition in which (i) *A* ascribes an *obligation* to *B* to do *x* (where it is assumed that *A* thinks the obligation is valid); and (ii) *A holds B* to the obligation. In order to count as holding another person to an obligation, one must care whether the other person fulfills it. In Frankfurt's (1988) terminology, this is a matter of investing oneself in the question of whether another person meets the obligation. This is

something more than merely staking something on the person's fulfilling the obligation. It is also different from endorsing the obligation, in the sense of considering it desirable or valuable that the person fulfill the obligation. It is an attitude of engagement toward the question whether she fulfills the obligation. For this reason part (ii) is separate from (i). It is possible to ascribe an obligation to somebody else without holding her to it.

The second main point about moral expectation (one we have already touched upon) is its connection to the appropriateness of blame. If one person has an obligation to do *x*, then it is appropriate for another person with the right kind of standing or authority to blame or punish her for not doing *x* (Wallace, 1994: 23; Mellema, 1998: 480; Paprzycka, 1999). Given that *A* ascribes an obligation to *B* to do *x*, this normally implies that *A* would blame *B*, or would find it appropriate to blame *B*, if *B* did not do *x* and had no excuse. (Often A has the standing to blame B. But if this were not the case, then A need only think that a person who did have the relevant standing would be within her rights to blame or punish *B* given nonperformance.) Thus, if in our earlier case, Robert is not disposed to blame Linda for nonperformance — and does not judge it appropriate that someone should do so — then according to the Moral Expectation account he does not trust Linda. This is why, unlike both the Staking account and the Predictive Expectation account, the Moral Expectation account explains the intuitive significance of a disposition to blame, or a belief that it is appropriate to blame, when one trusts. In a range of common examples of trust (a medic trusts enemy soldiers not to shoot at him as he brings medical supplies to the wounded; one's neighbor trusts one to look after a pet during an absence; a medical patient trusts his physician to do everything possible to diagnose his symptoms), the common core is a moral expectation that the other person will perform a certain way.

Certain other accounts place betrayal, rather than blame, at the center of the attitude of trust (Holton, 1994, p. 66, Hieronymi, 2008). This has some basis in our ways of speaking: "betraying a trust" is a standard idiom displaying the connection between acts of betrayal and trust relations. Betrayal carries a distinctive sentiment associated not just with blaming the person who betrays, but also with the idea that there has been a rupture of a personal relationship or agreement. Many cases of trust involve this personalized element, so that they are linked with the possibility of betrayal. Conversely, betrayal seems to presuppose trust (and so, the judgment of blame). However, not all cases of trust are such that feelings of betrayal are an appropriate response to unexcused nonperformance. For example, in purely professional relationships among disinterested colleagues where there are normative expectations of performance, trust and the possibility of blame can be found, but betrayal is not always appropriate. Even though the connection between trust and betrayal does indicate something about the moral character of trust, trust-with-the-possibility-of-betrayal is a more specific phenomenon than trust generally.[9]

The third main point about moral expectation is that it commits one to a certain kind of justificatory support for trust. As Wallace puts the point, "expectations should be capable of being supported by practical reasons" (1994, 22). If a person is obligated to do *x*, she must be *able* to do *x*, and have at least *adequate* justificatory reason to do *x*. Moreover, a person who attributes an obligation to somebody else ought to realize that this condition obtains, if she considers the matter. Therefore, when one person expects something from another, she should be sensitive to whether this second person can perform and has at least adequate reason to perform. Unfortunately, this requirement is often not met. For example, suppose I morally expect people not to get sick. I often feel blame towards others who are sick, and feel guilt about my own

sickness. Suppose that when I reflect upon these feelings I reject them because of the above requirement: I see that people cannot much control whether they get sick. But my feelings persist despite this reflection: in unreflective moments, I continue to experience these reactive attitudes toward sick people. In this case, I may still effectively hold others to my unreasonable expectation by exhibiting an attitude of blame toward them much of the time, even though I reflectively reject my attitude (Wallace, 1994: 43). Thus, moral expectations can be psychologically real without being reflectively endorsed, and sometimes after they have been reflectively rejected. This is all too often the case.

What kinds of reasons are plausibly adequate reasons for a trusted person to perform? It is not possible to launch a full discussion here, but here are some schematic thoughts. By having a moral expectation one person ascribes an obligation to another person. This obligation could be a standing obligation (e.g., a general obligation to help others in need, or not to interfere with their personal activities without reason); it could be an obligation related to their special roles or their relationships (e.g., the obligation of a professor to provide office hours to students); or it could be an obligation due to an agreement or practice that goes beyond either of the first two sources (e.g., an obligation to check on the neighbor's dog when one has done so many times before). I do not suppose this list is completely exhaustive, but it suggests some different putative obligations that might ground a moral expectation. Now one kind of performance-reason is simply an awareness of the obligation itself. But it is implausible to suppose that every moral expectation must carry with it the assumption that people have reason to act, as if by Kantian duty, from a recognition of the obligation itself. The reason for performance one supposes them to have may, instead, be of some other kind, such as the avoidance of blame, the

avoidance of punishment, the benefit they expect, the love they have for you, the advice they received from a respected source, etc. I take it that these reasons are compatible with the ascription of an obligation; they do not conflict with it (Nickel, 2007). Hence a moral expectation does not entail an implausible ascription of heavy moral motivations to those who are trusted.

So far, I have been developing an idea of moral expectation that may serve as a necessary condition on trust. But, together with the other partial conditions, is it jointly sufficient for trust? In the philosophical literature on this topic, an argument has been put forward that it is not. Consider the following case. Suppose a burglar observes Smith regularly visiting her elderly father in the hospital and robs Smith's house during one of these visits, while she is away. The burglar might ascribe an obligation to Smith to visit the hospital, and rely on Smith's acting on this obligation (Lahno, 2001: 178–82). Moreover, it is even possible that the burglar thinks the obligation that Smith visit her father in her hospital is (morally) sound, and holds her to it — that is to say, the burglar might care whether Smith visits her father, for the right sorts of reasons — even while relying on her to meet it for the sake of the burglary (Mullin, 2005: 324). In this kind of case, *A* engages in risky reliance on another person *B* to do *x*, and morally expects *B* to do *x*, but intuitively does not trust *B* to do *x*. This suggests that the conditions I have identified are not jointly sufficient for trust. It may also raise the worry that we have still not identified the core idea in an account of trust.

However, there are resources within my account to respond to this type of counterexample. For it appears that the action the burglar is relying on Smith to do is not the same thing that he morally expects Smith to do. The burglar relies on Smith to be out of the

house, say between the hours of 2:00 and 3:00 in the afternoon. Perhaps she predictively expects or bets that Smith will be out of the house during that time. But what the burglar *morally* expects Smith to do is not, except coincidentally, to be out of the house for those hours. Rather, in this peculiar example, what the burglar morally expects her to do is *to visit her father at the hospital*. If the burglar felt blame towards Smith for not visiting her father at the hospital, it would be because she didn't visit him at *any* time of day, not because she didn't visit him during those specific hours. If she visited him at a different time, the moral expectation would be equally fulfilled. Hence, all we need to do to respond to this type of objection is to specify that in a case of trust the thing the burglar relies on Smith to do must be the very object of his moral expectation.[10]

V. Rethinking Trust Empirically

My final aim in this paper is to suggest that the empirical phenomena of trust can usefully be seen as a matter of ascribing and holding people to moral obligations. The Moral Expectation view implies a distinctive view of common social phenomena examined under the heading of trust. For this reason, it is important to make a case that the Moral Expectation view has empirical application. For example, Hardin (2006: 16) complains of many accounts of trust that they are based on vernacular notions of trust that are explanatorily unfruitful. In what follows, I wish to address at least in a preliminary way the explanatory prospects of the Moral Expectation view. I will not be able to do full justice to the topic here, but I can suggest a few lines of inquiry.

First of all, I do not wish to suggest that *a priori* reflection about the concept of trust should carry the day against considerations of empirical and theoretical feasibility. There may be important reasons to adopt more concrete or "operational" vocabulary for some empirical investigations. For one thing, moral expectations may be too difficult to identify, psychologically, to serve as the basis of social scientific inquiry. It may be impossible or difficult in practice to determine what obligations people ascribe to others, and to distinguish moral expectations from other sorts of expectations they have and from their general dispositions of reliance. For these reasons, the account of trust I have offered here might not be the best way to articulate the concept of trust for rough-and-ready empirical purposes. However, even if social scientists elect to use a simpler, non-moral notion of trust for pragmatic reasons, it would be better to acknowledge this outright, rather than pretending that the concept has no moral contours.

Because it thinks of trust as holding only in those cases of willing, risky reliance where there is a moral expectation, the Moral Expectations view is narrower, extensionally, than the dominant views of trust. Is this extensional narrowing empirically significant? The kinds of disapprobation and sanctions that go along with moral expectation are very prevalent in human behavior, and people apply them in many contexts, so there is no concrete reason at the outset to think that this narrower conception of trust will exclude most of the phenomena that have already been studied empirically under the heading of trust. But existing empirical research does not raise this issue in any detail.

In explaining action in social situations, the Moral Expectations account of trust raises the possibility of a strong connection between the attitude of trusting somebody to do *x* and the

tendency of the trusted person to do *x* in order to meet the obligation ascribed as part of that trust — even if *x* is objectively not morally obligated, or even if it is objectively bad or objectively wrong.[11] Return to the case of Linda. Suppose Linda decides to set aside significant goals of her own, for example her schooling, in order to look after Robert's son. Suppose too that objectively she is not morally obligated to do this, because she has only minimal Good Samaritan obligations to Robert and his son. For example, it might be a permissible alternative for her to call the police or Social Services to report that Robert has abandoned his child. But suppose she isn't sure what her obligations are. She wants to be a good person — a trustworthy person — and she doesn't want to incur Robert's blame. (Note that if Linda can be sure that Robert won't blame her if she refuses to take care of the child, she has no such obligation-responsive incentives to do so.) In that case, there may be a significant disutility for Linda, and one that is objectively not morally required, in order to meet the moral expectations placed by Robert. In this way, misplaced trust can induce people to do what is bad for them or for others.

Consider an example involving patient trust in physicians. There is a phenomenon called "defensive medicine," discussed in the literature on physician response to the threat of medical malpractice, that appears to be linked to moral expectations. One type of defensive medicine, "assurance behavior" or "positive defensive medicine," is defined in one study as behavior that "involves supplying additional services of marginal or no medical value with the aim of reducing adverse outcomes, deterring patients from filing malpractice claims, or persuading the legal system that the standard of care is met" (Hiyama et al., 2006). Categories of assurance behavior in this study include ordering more tests than medically indicated, prescribing medications that are not medically indicated (such as antibiotics), referring patients to specialists unnecessarily,

and suggesting invasive procedures to confirm diagnosis. In this study, over 90% of a group of 131 Japanese gastroenterologists reported engaging in one of these forms of assurance behavior. The other type of defensive medicine is "avoidance behavior" or "negative defensive medicine": it is a matter of, for example, refusing to treat high-risk patients, relocating to a different state, avoiding high-risk subspecialties, retiring early, etc. This type of defensive medicine is sometimes blamed for the shortage of physicians in high-risk specialties and in rural areas (Mello et al., 2005).

It is often suggested that defensive medicine is a reaction to the threat of litigation. I wish to suggest that it can be seen at the same time as an indirect effect of misplaced trust. Suppose patients have moral expectations of their physicians, such that they hold their physicians to be obligated to provide them with referrals to specialists, to prescribe antibiotics and other drugs, and to initiate invasive diagnostic procedures, in a wider range of cases than the physicians actually should do these things. Patients' moral expectations imply a willingness to use the sanctions at their disposal, such as withdrawal as clients, medical malpractice suits, and reputational retaliation, to punish physicians who, despite patients' general attitude of charity, have not met these expectations. Physicians, picking up on these expectations as well as associated threats of blame and retaliation, then make efforts to meet patients' expectations or to withdraw preemptively from relationships that could lead to the worst consequences such as malpractice claims. As a consequence, they perform actions such as overprescribing antibiotics, overreferring to specialists, and overperforming invasive diagnostic procedures, as a consequence of patients' trusting them to do these things (Hall, 2004).

Studies of defensive medicine do not investigate, in a way that meets the highest standards of psychological inquiry, what the underlying motivation of the physicians is, nor is there a corresponding study of patients' moral expectations. But the interpretation I suggest is straightforward and subject to empirical inquiry. In the studies in question, it is assumed that physicians are reacting primarily to the threat of malpractice suits. But the threat of malpractice suits may be masking other forms of interpersonal motivation associated with the physician-patient relationship. Trust is thought to be central to such relationships (Davies and Rundall, 2000; Hall, 2005), so it would be natural to look to trust-based forms of motivation to explain positive and negative defensive medicine. In trusting physicians, patients place moral expectations on physicians. Some of these expectations are reasonable and some are not. But regardless of whether they are objectively reasonable, physicians may care about meeting them. Physicians, like the rest of us, are susceptible to the moral expectations of others in a way that can influence independent moral judgment. It is important to realize that misplaced trust can be bad, not because moral expectations are disappointed, but rather because they are *met*. This coincides with "positive" defensive medicine. When expectations are particularly high, another reaction is to avoid them, by withdrawing from the relationship that gives rise to them — as exhibited in "negative" defensive medicine.

In general, the idea that trust places a kind of pressure on the trusted person to behave in ways that are costly or risky to them is familiar. In fact, there are several overlapping kinds of pressure that being trusted places on the trusted person to perform, even in those cases where performance is costly for her or for others. I will mention several. First, it is often the case that the trusted person can foresee that there will be occasion for reciprocal reliance in the future. She

may reason that if she performs now, she can expect to be able reciprocally to trust the person who trusts her and gain future benefits. Hardin discusses this under the heading of what he calls "iterated prisoners dilemmas" (2006: 44–6). Second, the trusted person may wish to enjoy the good opinion of others, either because this is valuable to her in itself or because it will help her to earn a better reputation with others (Pettit, 1995). Third, the trusted person may wish to avoid a threat of punitive sanctions or retaliation. The fourth kind of pressure to perform, which I have focused on in this section, is a the trusted person's motive to meet the moral expectation to do *x*. This motive cannot be reduced to any or all of the other three. Caring about whether one meets the moral obligations ascribed by others is not the same as the desire for future reciprocal benefit, the desire for a better reputation, or the desire to avoid the consequences of a threat.

Given the powerful combination of these motives, it is surprising that so much of the literature on trust, particularly on patient-physician trust, assumes that the pitfalls of trust mainly occur when trust is disappointed (e.g., Davies and Rundall, 2000). With a bit more careful consideration, a very different kind of worry suddenly seems obvious: that these trust-based motives might cause people who care about being trusted to do something bad, i.e., costly, damaging, or unjust.

Notes

1. It falls outside the scope of this paper to defend my view of trust against prominent affective and virtue theoretical accounts of trust such as those offered in Jones (1996), Potter (2002); and Thomas (1990).

2. Although there are other ways of articulating the predictive notion of expectation, for example in terms of some other degree of likelihood besides .5, or as the highest-probability outcome in a ranking of salient possible outcomes, the points I make in what follows could be made *mutatis mutandis* against these alternative accounts. This is because, as I go on to argue, there is no principled lower limit to the trustor's ascribed likelihood of the trustee's performance, so long as the gain is sufficient to make it a worthwhile bet, and so long a moral expectation is adopted by the trustor.

3. Faulkner (2007) differentiates two kinds of trust, predictive and affective trust. Faulkner's notion of predictive trust is subject to the criticisms outlined in sections II and III. Affective trust, by contrast, is according to Faulkner based on a relationship in which the trusted person is responsive to the consideration that she is being relied upon. In my view, this condition is too strong because there are cases of trust in which the trusted person is not responsive to such considerations (Nickel, unpublished ms.).

4. Hardin's account of trust is highly distinctive. He characterizes the grounds for one's expectations in terms of what he calls "encapsulated interest" (Hardin, 2006: 17). However, this distinctive feature of Hardin's account will not matter to the criticism here, which is to the very idea of an Predictive Expectation account. In fact, the point about encapsulated interests could be divorced entirely from a Predictive Expectation account, by making encapsulated interests the ground for some judgment or mental state other than a predictive expectation. Thus my criticism of Hardin is *ad hominem* in the sense that it relates to an aspect of his theory that could give up without losing his most distinctive claim.

5. Coleman (1990: 100) attributes this sort of view to Deutsch (1962), and criticizes Deutsch's view on similar grounds to those I present below in section III.

6. This develops an argument that Coleman presents against Deutsch, using the example of trusting a confidence man (Coleman, 1990: 100). Herreros (2004: 8) considers this line of thought, but does not seem to think of it as an argument against a predictive view.

7. Moreover, it does not rule out the possibility that as an actual matter of fact we sometimes act against our own better judgment, even purposely. It does, however, hold that this renders our action difficult to explicate rationally and to justify; hence that we are under rational pressure to act as our sense of benefits and drawbacks dictates.

8. It is not part of the case that according to Robert, Linda knows (or believes) he is relying on her. On neither of the accounts we have considered, nor on my own positive account, is this a requirement for trusting.

9. The possibility of betrayal is linked with another phenomenon of trust, "evidence-resistance." It has sometimes been suggested that trust carries with it a special resistance to the thought that the trusted person has violated one's expectations or done something evil or stupid. One takes an attitude of charity toward the trusted person, exhibiting a reluctance to conclude that the person has culpably done wrong (Baker, 1987). This claim is applied by one writer to professional/client relationships as well (Hall, 2005). This phenomenon of "evidence-resistance," like betrayal itself, is not a universal feature of three-place trust. The reasons for being charitable in interpreting the behavior of another person may be quite different from the reasons for trusting her to do a specific thing. It seems to me that it is only by taking a charitable attitude toward the trusted person's behavior that one becomes susceptible to betrayal of one's trust.

10. Thanks to Michael Blome-Tillmann for discussion of this point.

11. Some philosophers even build this into their concept of trust, by holding that the trusting person must ascribe such reliance-dependence to the trusted person in order to count as trusting at all (Faulkner, 2007). This strikes me as an implausible condition.

References

Baier, A. (1986). Trust and Antitrust. *Ethics*, 96, 231–60.

Baker, J. (1987). Trust and Rationality. *Pacific Philosophical Quarterly*, 68, 1-13.

Barber, B. (1983). *The Logic and Limits of Trust*. New Brunswick, N.J.: Rutgers University Press.

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

Davies, H.T.O., and Rundall, T.G. (2000). Managing Patient Trust in Managed Care. *The Milbank Quarterly*, 78, (4), 609–624.

Deutsch, M. (1962). Cooperation and Trust: Some Theoretical Notes. In M.R. Jones (ed) *Nebraska Symposium on Motivation*. Lincoln: University of Nebraska Press, pp. 275–319.

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

Field, J. (2003). *Social Capital*. New York: Routledge.

Frankfurt, H. (1988). *The Importance of What We Care About*. New York: Cambridge University Press.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. London: Hamish Hamilton.

Gilbert, M. (1989). *On Social Facts*. Princeton, NJ: Princeton University Press.

Hall, M.A. (2004). Can You Trust a Doctor You Can't Sue? *Depaul Law Review*, 54, 303–14.

Hall, M.A. (2005). The Importance of Trust for Ethics, Law, and Public Policy. *Cambridge Quarterly of Healthcare Ethics*, 14, 156–167.

Hardin, R. (2006). Trust. Cambridge, UK: Polity.

Herreros, F. (2004). *The Problem of Forming Social Capital: Why Trust?* New York: Palgrave Macmillan.

Hertzberg, L. (1988). On the Attitude of Trust. *Inquiry*, 31, 307–322.

Hieronymi, P. (2008). The Reasons of Trust. *Australasian Journal of Philosophy*, 86, 213–236.

Hiyama, T., et al. (2006). Defensive Medicine Practices Among Gastroenterologists in Japan. *World Journal of Gastroenterology* 21, (12), 7671–7675.

Hobbes, T. (1968). *Leviathan*. New York: Penguin.

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

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

Keynes, J.M. (1921). *A Treatise on Probability*. London: Macmillan and Co.

Lahno, B. (2001). On the Emotional Character of Trust. *Ethical Theory and Moral Practice*, 4, 171–189.

Lewicki, R.J., McAllister, D.J., and Bies, R.J. (1998). Trust and Distrust: New Relationships and Realities. *Academy of Management Review*, 23, (3), 434–458.

Luhmann, N. (1979). *Trust and Power*. Davis. H., Raffan, J. and Rooney, K. (trans). New York: John Wiley and Sons.

McLeod, Carolyn. (2000). Our Attitude Towards the Motivation of Those We Trust. *Southern Journal of Philosophy*, 465–79.

Meijers, A. (2000). The Relational Ontology of Technical Artefacts. In P.A. Kroes and A.W.M. Meijers (eds.) *The Empirical Turn in the Philosophy of* Technology. Elsevier Science, Amsterdam, 81–96.

Mellema, G. (1998). Moral Expectation. *The Journal of Value Inquiry*, 32, 479–88.

Mello, M.M., et al. (2005). Effects of a Malpractice Crisis on Specialist Supply and Patient Access to Care. *Annals of Surgery*, 242, (5), 621–8.

Mullin, A. (2005). Trust, Social Norms, and Motherhood. *Journal of Social Philosophy*, 36, (3), 316–330.

Newton, K. (2001). Trust, Social Capital, Civil Society, and Democracy. *International Political Science Review*, 22, (2), 201–214.

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

—. (n.d.) Testimony, Trust and Normativity. Unpublished ms.

Paprzycka, K. (1999). Normative Expectations, Intentions, and Beliefs. *Southern Journal of Philosophy*, 37, (4), 629–52.

Pettit, P. (1995). The Cunning of Trust. *Philosophy and Public Affairs*, 24, (3), 202–225.

Potter, N.N. (2002). *How Can I Be Trusted?* Lanham, MD: Rowman and Littlefield.

Searle, J. (1995). *The Construction of Social Reality*. New York: Free Press.

Seligman, A.B. (1997). *The Problem of Trust*. Princeton, New Jersey: Princeton University Press.

Sen, A. (1982). *Choice, Welfare, and Preference*. Oxford: Blackwell

Smith, M. (2000). The Reality of Moral Expectations: A Note of Caution. *Philosophical Explorations*, 3, 232–8.

Strawson, P.F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48, 1–25.

Thomas, L. (1990). Trust, Affirmation, and Moral Character: A Critique of Kantian Morality. In O. Flanagan and A.O. Rorty (eds) *Identity, Character, and Morality: Essays in Moral Psychology*. Cambridge, MA: MIT Press, pp. 235–57.

Wallace, R.J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.