האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

TRUST, DISTRUST, AND IN BETWEEN

By

EDNA ULLMANN-MARGALIT

Discussion Paper # 269 Sep 2001

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Chapter 3

Trust, Distrust, and In Between

EDNA ULLMANN-MARGALIT

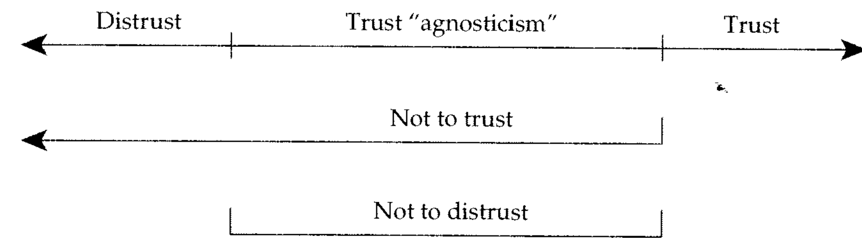
THE NOTION of trust has been the focus of intensive research in recent years. Given the negation relation between trust and distrust, a good understanding of distrust may be a useful way of shedding additional light, even if indirectly, on trust. In a similar vein, the attempts in psychoanalysis to understand the pathological mind have always been taken as contributing to a better understanding of normalcy, and a grasp of "politica negativa" as a necessary step on the way toward a more solid foundation of a positive political theory. If I want to know about the bright side of the moon, I may do well to look at its dark side, too (Margalit 2001, 127–28).

I approach distrust as a problem in practical reasoning, one that deals with the rules and strategies of action that we are to adopt in situations of social interaction in which the question of trust versus distrust comes up. This approach distinguishes itself from the subjective probability approach, which asks under what conditions we are to accept a hypothesis of distrust.

Normal linguistic use suggests the existence of an interim zone between clear cases of trust and clear cases of distrust. Trust and distrust, while mutually exclusive, are not mutually exhaustive. That is, I cannot both trust and distrust you, at least not with respect to one and the same matter (say, with respect to your writing a genuinely warm letter of recommendation for me), though it is entirely possible for me neither to trust nor to distrust you with respect to the same matter—or, indeed, in general. I may, in other words, be agnostic in the matter of trusting you; trust and distrust negate each other but do not complement each other (Hardin 2001, 496).

Still, if I distrust you, this surely means that I do not trust you. The converse, however, does not hold: if I do not trust you, I may actually distrust

Figure 3.1 A Trust-Distrust Continuum



Source: Author's compilation.

you, but this is not necessarily so. And what if I do not distrust you? Does this mean I trust you? Ordinary use would not quite accept that. This set of relationships is represented pictorially in figure 3.1. Clear cases of trust are on the right; clear cases of distrust are on the left. In between lies the spectrum of cases characterized by neither trust nor distrust. Everything to the left of the area marked as "trust" is the complement of trust, namely, "not to trust." As can be read off the diagram, the area of "not to trust" covers the area of "distrust" along with the no-man's-land of neither trust nor distrust (in the figure, "trust agnosticism"). That is, if I do not trust you, this could mean either that I distrust you—that is, that I have reasons to positively distrust you—or, more minimally, that I just have no reasons to trust you (nor to distrust you either). All of this accords, I believe, with our normal and intuitive linguistic use.

Think of driving on the highway. A good driver will do well not to be too trustful of the other drivers and to resort to so-called defensive driving. At the same time, there is no reason for her to distrust all other drivers altogether. After all, she and they share an interest in not colliding and in completing their respective journeys safely. This is a common situation in which people find themselves interacting in an impersonal manner with others and in which the question of trust does come up in some "thin" sense, as relating both to the motivations and to the competence of the other(s). While there is no room for trust here, there are normally no specific reasons for distrust, either.

As can be seen from this example, the complement of "distrust" is not symmetrical to the complement of trust. Had there been symmetry, the complement of "distrust" would comprise everything to the right of the area marked as "distrust"—including, in particular, the area marked as "trust." This would mean that if I say "I do not distrust you," I could plausibly be interpreted as saying that I actually trust you. This, I believe, does not accord with accepted use.

When I say that I do not distrust my secretary, I take it that you will understand me as saying that I do not have reasons to distrust the

secretary. I take it also that you will understand me, by implicature, as saying that neither do I have reasons to trust her. In other words, "not to distrust" is a narrower concept than a proper complement of "distrust" would be. "Not to distrust," then, is restricted to the middle section of the line in figure 3.1. It relates only to the no-man's-land area designated as "trust agnosticism," that is, to the area of neither trust nor distrust.

This asymmetry between the complementary notions of "not to trust" and "not to distrust" goes to the heart of the larger and much-discussed theme of asymmetries between trust and distrust. Against the backdrop of this larger theme, the bulk of the present study focuses on this middle ground, which comprises cases in which one has reasons neither to trust nor to distrust.

In the case of belief, too, there is a middle section, a no-man's-land, in which it is not the case that one believes that p nor is it the case that one believes that *not* p. However, this analogy indicates a connection between the notions of trust and belief that goes beyond the merely formal aspect. There exist, in fact, deep-level connections between the ideas of trust and belief. The Hebrew words for "trust," "belief," and "faith" all share the same three-letter root, "e-m-n." The faithful, that is, the believers, are the trustworthy, and to believe in God is tantamount to putting one's entire trust in him. When the Lord promised Abraham a son and heir, it is said of Abraham that "he believed in the Lord"—even though he knew that his wife Sarah was barren and beyond childbearing age. The traditional interpretation of this phrase is that Abraham trusted in the Lord. The Lord rewards him for this trust: "He counted it to him for righteousness" (Gen. 15:6; King James version).

Full Trust

The working analysis of trust presented here is an attempt to capture the conditions under which we would trust someone—or, rather, the conditions under which we would, with respect to a certain matter, trust someone in full. It relates primarily to the endpoint case, where I trust you fully about something. Note that the adjective "full" qualifies the degree of trust with respect to a given matter, not the range of matters with respect to which there is trust. When I say that he fully trusts his doctor I mean that he trusts her qua doctor to the fullest degree and without reservations. I do not mean by this that he trusts her about everything outside of their doctor-patient relationship.

I have good reason to fully trust you with respect to some matter when I believe that

1. you intend to behave or act in this matter so as to promote my interests and my general well-being;

2. you intend to promote my interests qua my interests (whether or not they coincide with your interests);
3. with respect to the matter at hand, you have the competence to behave or act so as to promote my interests.

In other words, my full trust in you requires that I attribute to you intention, right reason, and competence.

It is possible that the notion of full trust, the endpoint case, is not often applicable in practice. I start my account with it not because it is the most common case of trust but because it is a useful analytical strategy. Good grasp of the endpoint pure case gives one a handle on the range of deviations from it. The systematic possibilities of negating the notion of full trust provide us with the wide spectrum of cases that stretches from mere lack of trust to the opposing endpoint of full distrust.

Intention and Competence

Full trust, according to this analysis, is based in principle on reasons that include both intention and competence. These interact in subtle ways. They are not, in general, equal in weight: the competence component is secondary. If I have trust in your intentions—that is, if I have the required belief as regards your intentions or motives—but lack the required belief as regards your competence to act in a manner that will promote my interests, I would not say I distrust you. I may in fact still trust you. (Think of an incompetent mother: might not her children still trust her?) Not so in the converse case: if I have trust in your competence but not in your intentions, I would probably say that I distrust you. Note that when I say that I trust my surgeon, I may at times mean no more than that I think highly of his competence. That is, I assign a high probability to the surgeon's success in the operation. In such cases it will be more accurate to say that I *have confidence* in my surgeon, or that I *rely* on him, than that I trust him.[1]

In addition to the components of intention and competence, there is the additional "right-reason" component.[2] If I think someone intends to promote my interests but not for the reason that they are my interests, I do not necessarily distrust that person, though I cannot rightly say that I trust her, either. Suppose I hire a lawyer, whom I do not know personally, to represent my case strictly on the strength of what I have heard about her competence. In such a case, in which I assess that her chances of succeeding are good, I might say that I rely on her, or that I have confidence in her, or even that I have full confidence in her, but not that I fully trust her. Moreover, even if I believe that, being ambitious, she fully intends to win the case and thereby to promote my own interests, this still does not suffice to make it a case of full trust. Full trust, on the proposed analysis, requires that the lawyer's wish to promote my interests will be for the

right reason—namely, for the reason that these are my interests, and not, say, because she wants to become rich and famous. When I fully trust my lawyer with respect to this (or any other) case, I have good reason to believe that even in the case of a conflict of interests she will take my side.

Less Than Full Trust

Still, in intermediary cases between full trust and full distrust, it may at times be in order to speak of trust. I may trust my travel agent or my teaching assistant or my representative in Congress about a given matter to some degree lower than full trust. This will be so if the conditions in the proposed analysis are relaxed in some appropriate ways. For trust that is less than full trust, we may, for example, consider dropping the condition of right reason, or the condition of competence, or both. Alternatively, we may consider weakening each or both of them rather than dropping them altogether. In the case of full trust, both of these conditions involve *good* reasons, based on full-fledged beliefs. An obvious way to weaken them, therefore, would be to turn them into conditions that involve less compelling reasons: "I have some reason to believe that—," or "I have prima facie reason to believe that—" might do.

The intention condition can be weakened in a similar manner, that is, from my having good reason for my belief in your intentions to having some reason or prima facie reason for my belief in your intentions to promote my interests.[3] This weakening notwithstanding, I am assuming that the intention component goes to the core of the notion of trust and therefore that it cannot be dropped. In one version or another, it is indispensable. This is, in principle, what distinguishes the notion of trust from the notions of reliance or confidence.

To the extent that I am in a position to assign a (sufficiently high) probability to your acting in a way that will promote my interests, I may rely on you or have confidence in you (to the appropriate degree). But it is my having reasons to believe that you *intend* to act so as to promote *my* interests that makes me trust you. This assumption becomes problematic, however, when attention is shifted from trusting an individual to trusting an institution, such as the Supreme Court or a university. The question then comes up, how are we to construe the intentions of institutions? This problem suggests that either it is possible to relax the intention condition in further ways or we must construe talk about trusting institutions in different terms.

Symmetry, Mutuality, and Familiarity

The trust relation as so far portrayed is a vector in the sense that it has direction. It is, therefore, asymmetrical: it flows from me to you. All three conditions in the proposed analysis involve my beliefs about you and not

your beliefs about me. Not only does the notion of trust not emerge from the proposed analysis as a symmetrical relation, but experience shows that it is, in fact, often engendered within hierarchical relationships in which it is not—nor is it expected to be—reciprocated. Moreover, this is especially true in cases in which the hierarchical relations involve loyalty. When a secretary is known to the employer to be loyal to her, then the employer may develop trust in her secretary. Thus loyalty goes up in the hierarchy and trust goes down. This pattern characterizes such old-fashioned relations as those between kings and subjects, noblemen and vassals, and even husbands and wives in the old paradigm of marriage.

The account thus makes no assumption about mutuality. While my trustful relation toward you, as such, is asymmetrical, you may or may not trust me in return with respect to the same particular matter (or indeed with respect to additional matters). At the same time that it makes no assumption about mutuality, the trust relation emphatically does not *rule out* mutuality. Not only does it allow for mutuality, it may indeed expect it and call for it. Just consider the paradigmatic trustful relationship—a marriage of love, in which trust is assumed and expected to be symmetrical and mutual.

Note, as an aside, that the cluster of issues concerning symmetry, mutuality, and hierarchy extends in interesting ways to cases of relations of trust inside and outside of a group. Members of a family or clan may be loyal to one another and trust one another over a wide range of issues and distrust anyone from outside the family or the clan. (Consider the Mafia as an extreme case.) Moreover, a strong measure of distrust on a variety of issues on behalf of the members of some group—say, a minority—toward the members of another group—say, the majority—may create a strong bonding among the members of the first group. This bonding may even induce a relation of mutual trust among the membership of the minority group. In extreme cases, when the minority group resorts to violence against the majority, the phenomena of bonding and trust may eventually be engendered within the majority group itself (see Gambetta 1988; Banfield 1958).

The trust relation according to the proposed analysis does, however, make an assumption about mutual familiarity and acquaintance. It assumes some preexisting relation between the parties. My fully trusting you means that I believe that you intend to act so as to promote my own well-being and, moreover, that I believe that you intend to act in that way precisely in virtue of your wanting to promote my well-being. It is not reasonable that I should form these beliefs if I do not already know you, and you me, to some extent.

Trust relations can certainly flourish when the relationships between people qualify as thick: that is, when people are connected to one another by rich networks of family, clan, neighborhood, having grown up

together or gone to school together, or otherwise sharing a past (Margalit 2002; see also Williams 1988). But when we move on to relations of trust that involve less than full trust, thick relations are neither required nor assumed. Indeed, the force of the account is supposed to derive, in part, from its applicability to cases of casual acquaintances, that is, to cases that qualify as rather thin relations. Nonetheless, the point here is that they cannot be entirely thin: trust is, after all, a personal, not an impersonal, relation. This point again poses a problem for the institutional case. If it can be said that you trust the police, what sort of "mutual familiarity and acquaintance" between you and the police is one talking about here?

Trust Differentiated

The problem of trust is here presented within the framework of practical deliberation. When I assess my options for action, given the situation I am in, my decision as to how to act may at times depend on whether or not I have reasons to trust the person I am about to interact with. Thus framed, this is different from the theoretical problem of assigning subjective probability to the hypothesis that that person is trustworthy.

Still, the approach here offered is epistemic. The analysis of trust cites three belief conditions as reasons for my fully trusting someone. The beliefs involved are propositional: they have to do with "belief that," not with "belief in." One's faith, as expressed in one's belief in God, is related in religiously important ways to one's trust in God. But this is not the notion of trust I wish to focus on here. "Belief in" is, then, a notion from which our notion of trust has to be distanced.

Another notion from which trust has to be differentiated is confidence or reliance. The latter notions do not essentially involve the imputing of intentions; they lend themselves more readily to the subjective probability approach. I may rely on, or have confidence in, some*thing* (a bridge, for example), not only in some*one*. Trust, in contrast, relates only to people.

Finally, trust as here approached must be distanced from emotions. To be sure, when I trust you, and more so when I fully trust you, and especially when I fully trust you with respect to a wide range of issues, it is likely that there is warmth between us and that various other feelings and emotions may be involved. Trust-related emotions are worthy of analysis, but they must be kept separate from the account of trust offered here (Barbalet 1996).

Negating Trust

Trust and distrust, though they do not complement each other, do negate each other. How is trust negated? The proposed analysis addresses three conditions—intention, right reason, and competence—each of which

begins with a belief clause. A clause that begins with "I believe that p" can be negated weakly or strongly: compare "I do not believe that p" with "I believe that *not* p." The three conditions, all beginning with "I believe that," yield various combinations of these negations. As a result there are various negations of trust, at varying degrees of strength and located at different points in the space that opens up between full trust and full distrust.

Full Distrust

When I lack the belief that you intend to act in my best interests with respect to a given matter, I do not trust you. I begin to distrust you when I am in a position to form the actual belief that you do not intend to act in my best interests in that matter. My distrust in you increases when I become suspicious of your intentions, and it increases still further when I come to form the belief that you actually intend to act *against* my interests in the matter at hand. But this is not quite the extreme case yet. My belief that you intend to act against my interests may derive simply from my perception that our interests diverge and that you take your interests to trump mine. (When the difference between our interests is large, it might not even be seen as a violation of trust that you do what serves your interests but harms mine.) More generally, my belief that you intend to act against my interests may derive from the conjunction of my belief that you are concerned with promoting your own interests and not mine and my belief that your interests diverge from mine.

A more extreme case of distrust occurs when I believe that you intend to act against my interests with respect to the given matter, fully knowing that they are my interests. The most extreme case of distrust is encountered when I believe that you intend to act, with respect to that matter, against my interests qua my interests—that is, *because* they are *my* interests.[4] The extreme opposite of full trust is arrived at when, in each of the three conditions required of full trust, the expression "to promote my interests" is replaced with "to oppose (or harm or damage) my interests." Full distrust, then, also involves an intention component, a right-reason component, and a competence component.

I have good reason to fully distrust you when I believe that you intend to behave or act so as to harm my interests, with respect to a given matter, in virtue of their being my interests and that you have the competence to thus harm my interests. As in the case of full trust, the intention component and the competence component may interact in various ways, to different effects. Suppose that I believe that you want to harm my interests because they are my interests but that I also believe that you are generally powerless or incompetent. Here I would surely distrust you a lot, but at the same time this will be of little practical consequence,

and I will have little to protect myself against. Henry Fielding (1743/1964, 94) gives the intricate advice, "Never trust the man who hath reason to suspect that you know he hath injured you." This advice serves as a reminder that human relationships that lead to trust and distrust are more complex and multidimensional than the account here offered may lead us to believe. This also helps underline that full distrust, like full trust, is a personal relationship that cannot be founded on entirely thin relations.

Self-Trust and Self-Distrust

The notion of symmetry was invoked earlier; let us here briefly consider reflexivity. Can one trust, or distrust, oneself? "If you can trust yourself when all men doubt you / But make allowance for their doubting too," says Kipling (1910/1999, 496), making self-trust an item on his famous list of what's required for one to be a man. Of course, the use of "trust" here is loose, such that your trusting yourself is not entirely distinguishable from your being self-confident or self-reliant. Still, the proposed analysis of trust does extend quite naturally to the idea of self-trust. It makes sense for me to believe of myself that I am motivated to promote my own interests precisely qua my own interests and also, at times, that I believe that I am competent to do so. It thus makes sense to comment on someone that she trusts herself—or, indeed, in special circumstances, that she does not.

At the same time, the proposed analysis of trust does not extend itself to the idea of self-distrust, and quite appropriately so. When John Armstrong (1744/1804, 141), a U.S. army officer and the secretary of war from 1813 to 1814, says "Distrust yourself, and sleep well before you fight / 'Tis not too late to-morrow to be brave," he does not quite mean literal distrust: "do not be over-confident" is roughly what he means. The idea that I may be suspicious of my own motivations, or that I may want to hurt my interests because they are my interests, does not quite make sense—or, if it does, it belongs in the pathological department.

Samson and Delilah

Contrary to the impression that the account here offered might have created, human relationships of trust and distrust are highly complex and not always easy to disentangle. The biblical Samson did not trust his wife Delilah with the secret of his great strength. He had his reasons to doubt whether his own well-being was closest to her heart, and he had his reasons to suspect that her loyalty was rather to her kinsfolk and his bitter enemies, the Philistines. Did he actually distrust her? Delilah thought so, and she made him pay a price for his distrust. Any personal relationship that one expects to be based on trust goes sour when distrust creeps in. The disclosure of secrets is commonly taken to be a hallmark of trust, and giving secrets away as a hallmark of betrayal. Delilah kept pestering Samson to disclose to her the secret of his strength. Samson's repeated refusal to do so marred their marital relationship. "And it came to pass, when she pressed him daily with her words, and urged him, so that his soul was vexed unto death" (Judg. 16:16, King James version). Finally, perhaps to save his marriage, Samson decided to confide to his wife the secret of his strength. "If I be shaven then my strength will go from me, and I shall become weak, and be like any other man" (Judg. 16:17, King James version). The price he eventually paid for confiding his secret to his untrustworthy wife was, of course, immense. He paid for his misplaced trust with his life.

Did Samson's distrust of Delilah change to trust when he told her his secret? Her behavior toward him up to that point certainly gave him no good reason to trust her. Could he have *decided to trust* her, even if he did not have the "right beliefs"? My account of trust precludes this possibility. But it does allow for the possibility that one who lacks the required beliefs will still proceed to act *as if* he had them. The facts (or fictional facts) remain that Samson entrusted Delilah with his secret and that she duly proceeded to betray him by making him sleep upon her knees, shaving off the seven locks of his hair, and turning him over to the Philistines. Whether by telling Delilah his secret Samson proved that his distrust in her had changed to trust or whether he was only acting as if he trusted her, thereby acting out a death wish, remains a moot interpretative question. But the possibility of lacking the requisite beliefs yet proceeding to act as if one had them occupies center stage in the discussion that follows.

The Presumption of Distrust

Both trust and distrust require reasons. You will trust me, to whatever degree, if you have sufficient reasons for that degree of trust; the same holds for distrust. But what if the reasons you have are not sufficient, or you have no reasons either way?

A seemingly straightforward answer in such a case would be that you should neither trust nor distrust but wait until you have reasons to do one or the other. Often, however, in situations of practical deliberation, the need to act, and therefore the need to trust or distrust, is pressing, and one cannot afford to wait it out. In our social interactions, many of our decisions and actions depend to some degree on the extent to which we trust or distrust other people. If we are in a position neither to trust nor to distrust, because we lack the requisite beliefs (about the others' intentions and competence), we may have to resort to acting as if we had them; we may have to decide to act as though we trust or distrust.[5]

The situation, then, is this: you must act, and you must act now. How you act turns in an essential way on whether you trust or distrust me. But, so we suppose, you have no (sufficient) reasons either way. How you solve this particular problem on this particular occasion may depend on what is at stake and on your personality. You may decide to play it safe and act as if you distrust me, or you may decide to take a risk on my trustworthiness and act as though you do trust me. But when you come to face this sort of situations repeatedly, and when we want to generalize from you to people in general, you may realize that what you need is a *default rule* that will tell you which way to turn in the absence of adequate reasons.

The kind of default rule we are looking for is a presumption (Ullmann-Margalit 1983). The very possibility of a trust-distrust gap—that is, of being suspended between trust and distrust—paradigmatically suggests a role for a default rule in the form of a presumption. The presumption thus provides a link between the separate categories of belief and action: it tells you how to behave when you do not have the beliefs that would normally tell you how to behave. In this respect the situation is analogous to the case of the guilt-innocence gap in the setup of a criminal trial. Between "proven guilty" and "proven innocent" there will be cases of "proven neither guilty nor innocent." What is a judge in a criminal case to do when the time comes for a decision between conviction and acquittal but the balance of evidence leaves him or her suspended between the two? Two alternative presumptions could, in principle, provide judges with the default rule that they need: to consider the defendant guilty unless proved innocent (beyond reasonable doubt, say) or to consider the defendant innocent unless proved guilty (beyond reasonable doubt). We know which of the two our society chooses.

Regarding trust and distrust, one possible presumption would be this: In case of doubt, act as if you trust—unless or until you have (sufficient) specific reasons for distrust. Once you have such specific reasons in the concrete situation in which you find yourself, the presumption is rebutted. The alternative presumption would be this: In case of doubt, act as if you distrust—unless you have (sufficient) specific reasons for trust. This, too, is rebuttable. Of course, it is hard to know whether to act as if you trust or distrust in the abstract, without a sense of the consequences and the alternatives. Suppose I am in a situation in which I will die unless I take a risk on some stranger's trustworthiness (with respect to a given matter): I would best act as though I trust the stranger. On the other hand, suppose that a friendly reporter, or my dinner date, asks me some personal questions whose public disclosure might be embarrassing to me. It might be best to presume distrust here, or at least to act as though I do not trust. Some rapid

balancing of costs and benefits may be at work here, possibly with a bit of "maximin" thinking.

Still, in some of the literature on trust it sometimes seems to be taken for granted that "fairly generalized distrust might make sense in a way that generalized trust does not" and that, in abstraction from any specific context, suspicion and distrust are "inherently well grounded" (Hardin 2001, 500).[6] Cannot these statements be taken to constitute a sweeping recommendation for the presumption in favor of feigning distrust over the presumption of feigning trust? If so, how is this recommendation justified? After all, we all know that our world would be a much more pleasant place if it supported a general, contextless presumption in favor of trust. The issues involved need further exploration.

This exploration requires a shift of gear. The remainder of this chapter adopts a "game-theoretical," strategic approach that no longer resorts to the highly personalized notions of full trust or full distrust. General lack of beliefs that justify my trusting (or distrusting) you are assumed. Yet I take it that it is nevertheless possible for me to decide to act as if I had the requisite beliefs. This is the intended meaning of the phrase I use here, "adopting the trust (or distrust) strategy."

With these understandings, let us note that the presumption in favor of distrust is first and foremost justified on the ground that it is considered the safer of the two alternative presumptions. Consider the following rough calculation of best and worst scenarios. First, the case of trust: Acting as if you trust when trust is in fact reciprocated can lead to successful cooperation and hence to mutual benefit and potentially to significant gain. Acting as if you trust when your partner is untrustworthy and trust is not reciprocated inevitably involves disappointment. It often involves worse: being betrayed or exploited. It may lead to serious damage. Consider next the case of distrust: Acting as if you distrust when distrust is reciprocated leads to whatever gains you are able to achieve on your own. But what does acting as if you distrust, when your partner is trustworthy and does not reciprocate with distrust, lead to?

Here we may want to look at two different possibilities. One involves situations in which, when you adopt the distrust strategy, your gains are not affected by whether or not your distrust is reciprocated by your partner. Your gains remain the same regardless of whether your distrust is unilateral or reciprocated. This kind of case may in the long run breed lone distrusters who essentially expect nothing from their partners, individualists who "go it alone." They would be indifferent as to whether their partners trust them or not. The other possibility involves situations in which unilateral distrust does benefit the distrusters, at the expense of their trusting partners. In these situations the trusters are being exploited by their partners. Note that it is not two

psychological types that are differentiated here but rather two different types of situations.

"Soft" Distrust

Let us refer to the two possibilities just presented as "soft" distrust and "hard" distrust, respectively. The adjective "soft" is justified by the fact that the lone distruster exhibits mere *lack of trust*. Distrust here is benign: it does not cause harm but merely protects against harm (see Hardin 2001, 495–96). "Hard" distrust, in contrast, is exploitative of and harmful to one's partner; it involves betrayal of the partner's trust.

When payoff matrices are drawn for these two types of cases, it is easily, and unsurprisingly, revealed that cases of "hard" distrust have the structure of a prisoner's dilemma (PD) game. (For the matrices, and for further elaboration on the ideas presented in this section, see Ullmann-Margalit 2002.) In PD-structured situations, the noncooperative choice, which in the case at hand means acting as though one distrusts, dominates the cooperative choice of acting as though one trusts. Since the PD structure and its implications are well known, I focus on cases of "soft" distrust. Here too, as in the case of "hard" distrust, adopting the trust strategy can lift you high or make you fall. It can be disappointing; it is risk dominated. Adopting the distrust strategy, in contrast, is in this sort of case basically even. It leaves one on some in-between plateau that is insensitive to changes in one's environment—as well as insensitive to the disappointments one may cause to others. One neither exploits possibilities nor exposes oneself to being exploited by others.

To the extent that "playing it safe" means hedging your bets, minimizing your potential losses, being risk averse, to act as though you distrust thus seems to be the safer choice not only in the cases of hard distrust but in the case of soft distrust, as well. But suppose rather than a one-round encounter the partners are in a situation that repeats itself. If the partners start by playing it safe and adopting the distrust strategy, they will remain stuck with a suboptimal equilibrium in future repetitions of the situation (see Hardin 2002, chap. 5). If, however, they succeed in coordinating on acting as if they trust, whether through communicating with each other or somehow independently, then both will reap the fruits of their cooperation. Neither of them will be tempted to "defect" to the distrust strategy in future repetitions of the situation.

If the situation is further generalized, not just from a one-round to a repeated situation but also from two participants to a community, is the presumption of distrust justified? This may be conceptualized as a Wild West community of rugged individualists. They are honest folks who are used to relying on no one and to exploiting no one. Still, in the long run, they need be neither blind nor averse to the possibility of trustful

cooperation and to its mutual benefits. In such a community the argument of "playing it safe" does not justify the adoption of the presumption in favor of distrust.

Hobbes's State of Nature

When people think of paradigmatic cases of having to choose between acting as if they trust and acting as if they distrust, it is PD-structured situations that they commonly have in mind—namely, cases of hard distrust. Many people seem to take it for granted that the distrust strategy dominates the trust strategy paradigmatically. To act as though you trust when you lack the requisite beliefs seems much worse than simply to take a chance: it seems as though it actually *means* to be a sucker, to expose yourself to exploitation.

The general outlook of hard distrust may well derive from the powerful hold that Thomas Hobbes's grim picture has over us, the picture of the state of nature as a state of suspicion of all in all and of a war of all against all. Indeed, Hobbesians often tend to interpret social interactions, whether on the micro- or macro-level, as one-round games. To the extent that these are PD-structured situations, this outlook tends to justify a general presumption in favor of feigning distrust.

But even Hobbes himself, in presenting what he calls the "precept, or general rule of Reason," distinguishes between two situations. Hobbes (1651/1968, 190) says that "every man ought to endeavour Peace, as farre as he has hope of obtaining it and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre." The first part of this precept contains what is for Hobbes the fundamental law of nature: to seek peace and follow it. The second part sums up what he refers to as the right of nature: to defend ourselves by any and by all means we can. There is nothing far-fetched or strained, as far as I can see, in interpreting the first part of the precept as applying to situations of soft or mild distrust and the second as applying to situations of hard or harsh distrust.

True, Hobbes did not believe that endeavoring peace in the hope of obtaining it would get one very far. A close reading of the relevant passages reveals how deeply convinced he was that we are doomed constantly to seek and use the advantages of war to defend ourselves. But the important point is that he did seem to recognize the possibility that the state of nature might be construed in terms of soft distrust as well as hard distrust.

The fact that not every situation of distrust is structured as a prisoner's dilemma is crucial here. As in those cases that fall under the category of soft distrust, mutual trust not only leads to a jointly beneficial outcome but it is a stable equilibrium, and it is accessible to the participants. This in itself suffices to cast serious doubt on the idea that

the presumption in favor of distrust should be considered a universal "default" presumption.

Further Observations

To be sure, trust is fragile. As soon as breach of trust occurs, for whatever reason and by however small a number of people, a tipping-point phenomenon is likely to occur, and distrust will rapidly prevail. Like Humpty Dumpty, trust, once shattered, may be beyond repair. But to the extent that situations of soft distrust exist and are recognized as such, the precariousness—as well as the preciousness—of trust in such situations may at the same time be recognized, too. Furthermore, it is not impossible to imagine situations in which, because I decide to act as if I trust you, I eventually bring it about that you do become trustworthy and deserving of my future trust. This is what happened with Victor Hugo's priest, who chose to take the risk of acting as though he trusted Jean Valjean, thereby making him trustworthy (Hugo 1862/1992). Perhaps this is what Samson hoped against hope would happen with Delilah.

As small children we have to start out with something like an instinctive conclusive presumption in favor of sweeping trust in the adults who care for us. After all, it is hard to imagine how small children could generally get on, let alone learn a language, if they started out with an instinctive attitude of suspicion and distrust. If this is crudely true, then a case can be made that, at least developmentally, it is the trust strategy that is for many the default strategy and distrust is learned. So perhaps distrust is not a foregone social conclusion on this consideration, as well.

In thinking about the default presumption it must be noted, finally, that not only do people divide empirically into instinctive trust presumers and instinctive distrust presumers: contexts divide, too, in paradigmatic ways. We naturally catalogue situations involving economic transactions, for example, as ones in which a presumption in favor of feigning distrust is justified. Familial and communal situations, in contrast, we are often quite happy to approach with a presumption of trust. When we go abroad or are otherwise outside of our habitual contexts, we are typically in doubt. But even in such cases it is too crude to counsel in favor of acting as though we distrust. We are often able to use various social cues, sometimes quite subtle ones, to sort out different contexts and to identify those that justify taking the risk of trust. Who of us has not encountered the classical yet puzzling case on the beach, when the total strangers who happen to sit next to us ask us to "keep an eye" on their belongings until they return from their dip in the water?

It may thus be the *content* of the situation in which we find ourselves, in its wider social context, that will argue against a pessimistic and suspicious adoption of the distrust presumption. Alternatively, it may be

the game-theoretical *structure* of the situation that will achieve the same purpose, once the distinction between soft and hard distrust is internalized and correctly applied. For one reason or another, it may, after all, be the case that distrust shall have no dominion.

Institutional Trust and Distrust

There is an impressive volume of social-psychological literature about trust and distrust within organizations. The bases of trust, the benefits of trust, and the barriers to trust have been studied extensively (a useful survey of this literature is given in Kramer 1999). Quite separate from this body of research that is concerned with the antecedents and consequences of trust and distrust within organizations, political theorists are also concerned with institutional trust or distrust. Their concern, however, is with the question to what extent the public displays trust or distrust toward this or that social institution, and with what implications to the polity.

It is taken to be a necessary condition of a well-functioning democracy that its citizens trust its institutions. In a sense—a somewhat ironic sense—social institutions are sometimes seen as trust mediators. On the one hand, there is the fact that in modern mass democracies, in contrast to the intimate city-states of ancient times, no level of general interpersonal familiarity and trust can be assumed. An important role of institutions, then, is to facilitate social transactions by essentially replacing the need for personal trust among citizens (see Hardin 2001, 518; also see Hardin 2002, chaps. 7, 8)—consider, for example, the role of legally binding contracts as a replacement for promises. On the other hand, there is the further consideration that once the institutions are in place, in order for them to fulfill their role as trust replacers, it is often supposed that citizens need to trust *them*. In terms of the example just cited, in order for contracts to work it is commonly said that people need to trust their country's legal system and its enforcement mechanisms.

A number of writers seem to diagnose a malaise in many contemporary democracies, which they believe to exhibit a general decline in institutional trust. This relates to both public and private institutions. There is substantial evidence, for example, that institutional trust in the United States has been declining for several decades—in federal government, universities, medical institutions, and journalism as well as in several major private companies (see Coleman 1990; Carnevale 1995; Ney, Zelikov, and King 1997).[7] These finding are alarming if the ability of institutions to function properly depends in no small measure upon public trust in them.

According to another view, representative democracy and distrust go together. "A certain amount of distrust," says Russell Hardin (2001,

517), "may be useful to a society or government. Certainly, large, modern democracies work better if we can be sure that there are professional distrusters or cynics or skeptics, people who act as watchdogs, raise alarms, or provide contrary information."[8]

Much can be said in an attempt to explore these two views and possibly reconcile them. How threatening—or healthy—to a democracy are various degrees of institutional distrust? Is the sort of distrust one is talking about when arguing that it is threatening quite the same as the sort of distrust one is talking about when arguing that it is healthy? Moreover, there may be interest in following the further body of research that tries to advance explanations, from a variety of perspectives, for the sources of the erosion in public trust in institutions in various countries. But the question I pursue is a different and, in a sense, a prior one. What does it mean to trust or distrust an institution?

Institutional Trust

As it stands, the proposed account of trust and distrust will not do for the institutional case. The analysis requires that for me to trust X, I need to entertain certain beliefs about X's intentions and about what motivates those intentions. Since it is to persons, not institutions, that we attribute such intentions and motivations, it would seem to be the case that X can only be a person and not an institution. If one accepts this line of thinking, it follows that our common, everyday talk of trust or distrust in institutions may have to be rethought and possibly revised.

Still, in an attempt to make sense of talk about trust or distrust in institutions, several ways may seem to be available to go around this obstacle. In principle, one may either see one's way to attributing intentions to institutions, if not directly then somehow indirectly, or one may see one's way to weakening the intention component or dropping it altogether from the analysis of distrust.

One may acknowledge, for example, that even though we attribute intentions primarily to individuals, we can nevertheless attribute intentions to institutions in some derivative or secondary sense. This line of thought puts the onus of the argument on clarifying the derivative sense in which it may be coherent to talk about the intentions of an institution. One way may be to argue that it is often the person who is the figurehead of an institution that embodies for us the institution as a whole. Roderick Kramer cites a speculation that "people may use the behavior of institutional leaders as reference points for gauging their basic beliefs . . . when appraising the trustworthiness of society's institutions in general. In other words, people may draw general inferences about institutional trust from the behavior of highly visible role models" (Kramer 1999, 589). Insofar as this is so, the question of trust in the police, say, or in the

Supreme Court is translated into people's beliefs about the intentions and motivations of the commissioner of police or of the chief justice.

A different route is to tinker with the intention component of the analysis of trust. Given that I trust (or distrust) you, is it really necessary that I entertain beliefs about how you intend to act and about what motivates your intentions? Is it not enough perhaps that I entertain beliefs and assess probabilities about how you are actually going to act?

I believe that this is not enough. The attribution of intentions is of the essence, so far as trust is concerned. Entertaining beliefs and probabilities about the future course of action of a person or of an institution has to do with the notions of reliance and confidence but not with trust. I believe, indeed, that as far as trust is concerned, talk of trusting an institution is misplaced. To say that we trust an institution is to be construed, rather, in terms of our reliance on an institution or of our degree of confidence in its competence and performance. This can be expressed, for example, in the probability we assign to its achieving its set goals—provided its goals accord with our interests. (I can also have confident expectations that an institution will achieve its goal and *therefore* distrust it because its goal is against my interest.)

More specifically and more crucially, talk of trusting an institution ought to be construed in terms of our degree of confidence that the institution will continue to pursue its set goals and to achieve them regardless of who staffs the institution. There is a *principle of substitutability* at work here: whenever the idea of substitutability comes up, the question to ask is what remains constant under the substitution. In the case at hand, when we express trust in an institution we express our belief that, even if the present officeholders in that institution were to be replaced with others, the performance of the institution would remain constant. In other words, so-called trust in an institution is tantamount to a belief in the *impersonality* of its performance, in addition to the belief that its goals are compatible with our interests. Given our account of trust, it is precisely this impersonality that prevents this attitude from counting as trust.

When we trust an individual, we expect his or her attitude toward us to be entirely personal. When we say we trust an institution, we expect its attitude to us to be impersonal. Can it be the same notion of trust that is invoked in both cases? I think not. Strictly speaking, in the institutional case it is a misnomer to talk of trust.

This may have to be somewhat qualified, though. Often, talk of trust is bound up with social role. I may trust my dean; you may trust the U.S. president; he may trust the federal court. What is meant here is something not entirely impersonal yet less personal than in the noninstitutional case.[9] In trusting the dean, I trust that she will not be corrupt, that she will not play favorites, that the interests of the institution will have priority at her heart, that she is competent, that she will try to do her

best. These express confident expectations, not trust in the full personal sense. Nevertheless, it must be acknowledged that talk of trust here is consistent with ordinary usage and cannot easily be dispensed with.

Institutional Distrust

The case of distrust in institutions, however, is different. Here, I believe that we do attribute intentions and motivations, and not just to the figureheads of the institution.

Consider, for example, the case of the ultraorthodox in Israel, who in recent years have expressed growing distrust in the Israeli Supreme Court. Consider also the case of the Arab Israeli minority, whose members now talk of having lost whatever trust they had in the Israeli police.[10] In expressing their distrust in the respective institutions, these people are conveying something other than a mere factual prediction to the effect that the Court or the police will act in ways that will not further their interests but will rather collide with them. Their expression of distrust has a surplus element that goes beyond expressing nonreliance or low degrees of confidence. Rather, I believe, these communities want to be understood as imputing intentions, diffusely, to those who staff the respective institutions.

What intentions can these be, given that there is no personal acquaintance and there are no personal relations between the individuals involved? At bottom, I suggest, the question of distrust in an institution boils down to one's belief in the unfairness of the institution—and to the ancillary belief that the unfairness works against one's interests. When an institution faces a crisis of trust, which is at the same time a crisis of legitimacy, this means that segments of the populace in need of recourse to the institution in question suspect it of operating in an unfair manner, a manner that goes against their interests. More specifically, in many cases this means that these members of the public tend to impute discriminatory intentions quite generally to the officeholders at all levels of hierarchy in the institution—for example, to all the judges or to all the policemen and policewomen. The discriminatory intentions may be racist, sexist, homophobic, anti-Semitic, antireligious, or what have you.

The flip side of any discriminatory intentions that make some people distrust an institution is that the very same institution may become highly trustworthy to people with other, opposing, interests. In a city where the police favor the Mafia, it may be expected that the general citizenry will distrust the police. Can we say that the members of the Mafia trust the police? Well, they sure do, in some sense. But their trust in the police is in the personal sense of trust, not in the institutional sense that is premised on impersonality and substitutability.

The mafiosi's trust in the police cannot be the trust we are after when we reflect upon the role of institutional trust in a healthy democracy.

Their trust is a perversion of the trust in institutions that is claimed to be required for mass democracy to work. A necessary condition for institutional trust worthy of its name is confidence in the fairness and impartiality of the institution. (It is not a sufficient condition, though, as the element of competence is missing.) When this condition is fulfilled, there is no imputing of personal intentions to those who staff the institution; the principle of substitutability holds.

The point, then, is that in contrast to the case of institutional trust, institutional distrust does involve the imputing of intentions. It involves a shared belief among groups of citizens about the personal intentions of the officeholders of the institution. These intentions are taken to be operative while the officeholders are executing the duties of their office. The typical belief here is that these officeholders are infected with discriminatory intentions against the members of the relevant groups and that these intentions result quite generally in unfair practices. The unfair practices are believed to operate in principle against the interests of those groups of citizens.

Another possibility for institutional distrust occurs when there is widespread belief that the institution is corrupt. David Hume (1748/1987) suggests that institutions should be designed in such a way that they would work well even if, in his well-known phrase, they were staffed by knaves. Should the design fail, however, or should the level of corruption of the knaves pass a certain threshold, the institution qua institution may be perceived to be corrupt. Here, too, the attitude toward the institution turns in an essential way on people's beliefs about the personal motives of officeholders of the institution at its various hierarchical levels. Once their personal motives become suspect—as, for example, when there is a shared belief that they serve foreign interests or are open to bribes—then general distrust in the institution qua institution reigns.

There may be an interesting difference between cases in which institutional distrust is based on partiality and those in which it is based on corruption. The first tend to be cases of *group* distrust, based on membership in groups defined by race, gender, ethnic origin, sexual orientation, and so forth. The second tend to be cases of *class* distrust, in which the institution is taken to operate in such a way that the rich can get away with things that the poor cannot. The two kinds of institutional distrust may of course overlap, and there may be various intermediate cases, too.

Conclusion

The account of trust offered in this chapter is a belief-based account: roughly, I trust you when I believe that you have the right intentions toward me. Trust can be differentiated from the related but importantly different notions of reliance and confidence. The analysis also affords

some insights into questions worthy of future attention, such as why trust is not, in general, a symmetrical relation and why trust can be reflexive but distrust not (that is, why it makes sense to say that I trust myself but not that I distrust myself).

The problem of trust as presented in this chapter is a problem of practical deliberation: how is one to act in a situation in which trust (or distrust) is required but the requisite beliefs are lacking? This is a problem because trust and distrust are exclusive but are not exhaustive: the absence of reasons to trust does not entail distrust, and the absence of reasons to distrust does not entail trust. Regarding those situations in which one has reasons neither for trust nor for distrust, can there be a reasoned policy in favor of acting as though one had reasons for either? The commonly held idea that the presumption in favor of acting as though you distrust is better and safer than the opposite presumption is probed. On the basis of the notion of soft distrust, this chapter argues that a pessimistic and suspicious adoption of the distrust presumption as a general rule of behavior is unfounded.

The account of trust in relation to institutions requires some modifications. Because institutions cannot have intentions in anything like the way persons have intentions, to say that we trust an institution cannot involve an ordinary ascription of intentions. What we ordinarily mean by trusting an institution should be construed not in terms of trust but rather in terms of our confident prediction that the institution will pursue its goals. Moreover, when we say that we trust an institution we expect the institution to be impersonal, whereas in trusting an individual we expect his or her attitude toward us to be entirely personal. In light of such considerations, in the institutional case it is a misnomer to talk of trust. Not so, however, in the case of distrust. Distrusting an institution is not a matter of confident predictions, and it does involve the assigning of intentions. Institutional distrust embodies one's belief that the intentions of the officeholders of the institution are discriminatory and that the institution is consequently unfair in ways that work against one's interests.

I owe gratitude to Russell Hardin, Cass Sunstein, and Avishai Margalit for discussing this chapter with me and for their valuable comments and ideas.

Notes

1. I learned about the role of competence in trust from Sidney Morgenbesser.
2. It is mostly this condition that distinguishes my account from Hardin's encapsulated-interest account of trust (elaborated in Hardin 2002). Strictly speaking, intention is subsumed under right reasons. The focus of the two conditions, however, is different.

3. The intention condition can also be strengthened. It may be the case that I believe that you intend to behave or act so as to promote my interests and my general well-being not only with respect to a particular matter but in all matters. This may be because I believe that you love me, as a parent or a spouse, or because I believe that you value me highly as a friend. We may say that in such cases my trust in you is not only *full* but also *complete*. That is, it relates both to the full degree to which I trust you with respect to any given matter and to the complete range of matters with respect to which I fully trust you: I trust you with everything and anything. Cases of full and complete trust may be rare, but they are not nonexistent. (See also note 4 for the analogous notion of *complete* distrust.)
4. Note that one may speak also of *complete* distrust, in analogy to complete trust (see note 3). My distrust is complete when it is not relativized to a particular matter, that is, when I distrust you with respect to everything and anything. This may be because you are my bitter enemy and I believe that you thoroughly hate me.
5. For a relevant discussion of an analogous distinction, between believing a proposition true and deciding to behave as if we believe it true, see Ullmann-Margalit and Margalit (1992).
6. While these are quotes from Hardin's article, they do not express his view.
7. See also Slovic (1993) for a discussion of distrust in nationwide institutions responsible for risk management in connection with technological hazards. It would surely come as no surprise if the Enron-Anderson debacles of late were shown to have produced serious new waves of distrust in financial institutions.
8. "Although often portrayed in the popular press and social science literature largely in negative terms, distrust and suspicion may constitute appropriate and even highly adaptive stances toward institutions. Vigilance and weariness about institutions, some have argued, constitute essential components of healthy and resilient organizations and societies. From this perspective, distrust and suspicion may, in a fundamental sense, constitute potent and important forms of social capital" (Kramer 1999, 590.) See also Ely (1980) and Warren (1999).
9. When I trust my lawyer qua my lawyer, this is less than full personal trust in a friend qua friend but more than just full confidence in *a* lawyer, as distinct from *my* lawyer. (Compare the lawyer example under the section "Full Trust.")
10. The acute crisis in Israel goes back to early October 2000, when the Israeli police shot to death thirteen Arab Israeli citizens during demonstrations that erupted in connection with the Palestinian uprising.

References

Armstrong, John. 1744/1804. *The Art of Preserving Health.* "Prefixed," a critical essay on the poem, by John Aikin. Philadelphia: Printed for Benjamin Johnson.

Banfield, Edward. 1958. *The Moral Basis of a Backward Society*. New York: Free Press.

Barbalet, Jack M. 1996. "Social Emotions: Confidence, Trust and Loyalty." *International Journal of Sociology and Social Policy* 16: 75–96.

Carnevale, David J. 1995. *Trustworthy Government: Leadership and Management Strategies for Building Trust and High Performance*. San Francisco: Jossey-Bass.

Coleman, James. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.

Ely, John Hart. 1980. *Democracy and Distrust*. Cambridge, Mass.: Harvard University Press.

Fielding, Henry. 1743/1964. *Jonathan Wild: The Journal of a Voyage to Lisbon*. Introduction by A. R. Humphreys. London: J. M. Dent & Sons Ltd.

Gambetta, Diego. 1988. "Mafia: The Price of Distrust." In *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta. New York: Basil Blackwell.

Hardin, Russell. 2001. "Distrust." *Boston University Law Review* 81(3, June): 495–522.

———. 2002. *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Hobbes, Thomas. 1651/1968. *Leviathan*, edited by C. B. Macpherson. London: Penguin.

Hugo, Victor. 1862/1992. *Les Miserables*. Translated by Charles E. Wilbour. New York: The Modern Library.

Hume, David. 1748/1987. "Of the Independency of Parliament." In *David Hume: Essays Moral, Political, and Literary*, rev. ed., edited by Eugene F. Miller. Indianapolis, Ind.: Liberty Fund.

Kipling, Rudyard. 1910/1999. "If." In *Rudyard Kipling*, edited and with an introduction by Daniel Karlin. Oxford: Oxford University Press.

Kramer, Roderick M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50: 569–98.

Margalit, Avishai. 2001. "Recognition." *Aristotelian Society* Suppl. 75: 127–39.

———. 2002. *The Ethics of Memory*. Cambridge, Mass.: Harvard University Press.

Ney, Joseph S., Philip D. Zelikov, and David C. King. 1997. *Why People Don't Trust Government*. Cambridge, Mass.: Harvard University Press.

Slovic, Paul. 1993. "Perceived Risk, Trust and Democracy." *Risk Analysis* 13(6): 675–82.

Ullmann-Margalit, Edna. 1983, "On Presumption." *Journal of Philosophy* 80(3): 143–63.

———. 2002. "Trust out of Distrust." *Journal of Philosophy* 99(10): 532–48.

Ullmann-Margalit, Edna, and Avishai Margalit. 1992. "Holding True and Holding As True." *Synthese* 92(2): 167–87.

Warren, Mark E. 1999. "Democratic Theory and Trust." In *Democracy and Trust*, edited by Mark E. Warren. Cambridge, U.K.: Cambridge University Press.

Williams, Bernard. 1988. "Formal Structures and Social Reality." In *Trust: Making and Breaking Cooperative Relations*, edited by Diego Gambetta. New York: Basil Blackwell.