Chapter 1

Trust

Usually, to say that I trust you in some context simply means that I think you will be trustworthy toward me in that context. Hence to ask any question about trust is implicitly to ask about the reasons for thinking the relevant party to be trustworthy. In chapter 2, I canvass some of the potentially many reasons for thinking someone trustworthy. One of the most important and commonplace is *trust as encapsulated interest*, which I discuss in this chapter. On this account, I trust you because I think it is in your interest to take my interests in the relevant matter seriously in the following sense: You value the continuation of our relationship, and you therefore have your own interests in taking my interests into account. That is, you encapsulate my interests in your own interests. My interests might come into conflict with other interests you have and that trump mine, and you might therefore not actually act in ways that fit my interests. Nevertheless, you at least have some interest in doing so.

There are two compelling reasons for taking up trust as encapsulated interest. First, such trust fits a centrally important class of all trust relationships. Second, it allows us to draw systematic implications for trust relationships across varied contexts, as subsequent chapters should make clear.

To begin, consider an example of trust from Dostoyevsky's *The Brothers Karamazov*. The example is instructive because it involves minimal conditions for trust as encapsulated interest: the only reason for trustworthiness is the incentive to sustain the relationship—in this instance, purely for its profitable character and not for any richer reasons. For many trust relationships there are additional considerations beyond monetary benefits. For example, I value my relationship with you in many ways, including for the favors we do each other, for the pleasures of talking and being with you, and for the mutual supports we give each other.

In *The Brothers Karamazov*, Dmitry Karamazov tells the story of a

lieutenant colonel who, as commander of a unit far from Moscow, has managed substantial sums of money on behalf of the army. Immediately after each periodic audit of his books, he takes the available funds to the merchant Trifonov, who soon returns them with a gift. In effect, both the lieutenant colonel and Trifonov have benefited from funds that would otherwise have lain idle, producing no benefit for anyone. Because it was highly irregular, theirs was a secret exchange that depended wholly on personal trustworthiness not backed by the law of contracts. When the day comes that the lieutenant colonel is abruptly to be replaced in his command, he asks Trifonov to return the last sum, 4,500 rubles, loaned to him.

Trifonov replies, "I've never received any money from you, and couldn't possibly have received any" (Dostoyevsky 1982 [1880], 129). Trifonov's "couldn't possibly" is an elegant touch because it drives home in a subtle way that the entire series of transactions has been criminal. The lieutenant colonel could not have wanted any of it to become public—not even at the cost of the final 4,500 rubles to keep it secret. Had it become public, he most likely would have gone to jail, and his daughter's marriage prospects would have been ruined. Unlike many of our important relationships, this one was partially abstracted from social context rather than being heavily embedded in a context that could govern the interaction and enforce trustworthiness. Trifonov's misappropriated rubles thereafter thread their complex way through Dostoyevsky's entire novel, wrecking lives while motivating the plot.

While their relationship was ongoing, Trifonov and the lieutenant colonel could each end their cooperation at any moment. Trifonov could have cheated at any time along the way, but then he might have lost more on forgone future interactions than he would have gained on his single, cheating defection, and he would have cheated a very powerful man. This was true so long as the interaction was expected to continue indefinitely. Once the interaction was to end, however, the incentive to the next mover was clearly to withdraw from the cooperation. It was the lieutenant colonel's misfortune that Trifonov was the next mover at the end and, at that time, the lieutenant colonel had lost his power.

Dmitry Karamazov says that the lieutenant colonel implicitly trusted Trifonov. After his sad day of reckoning, the lieutenant colonel would presumably have said that Trifonov was not trustworthy. Unfortunately, Trifonov was trustworthy just so long as there was some longer-run incentive for him to be reliable in their mutually beneficial relationship. The moment there ceased to be any expectation of further gains from his relationship with the lieutenant colonel, Trifonov had no incentive to be trustworthy in this highly irregular commercial

dealing between de facto crooks. Not surprisingly, he ceased to be trustworthy. The lieutenant colonel might have held the contradictory hope that Trifonov the crook was a man of honor who was, in that sense, trustworthy. If so, then he clearly misjudged his man.

Ordinary trust between individuals often fits the tale of Dostoyevsky's 4,500 rubles as a minimal condition and, more generally, the encapsulated-interest model as defined in this chapter. Understanding that model of trust can help us to understand many other issues. The encapsulated-interest conception fits three distinct categories of interaction: interactions that fit the model of the iterated one-way trust game, iterated exchange interactions as modeled by the prisoner's dilemma, and interactions in thick relationships. These categories of trust relations represent increasing complexity of the interactions. In all of these interactions, *trust is relational*. That is to say, my trust of you depends on our relationship, either directly through our own ongoing interaction or indirectly through intermediaries and reputational effects. If we have no or only a passing relationship, we are not in a trusting relationship.

Trust as Encapsulated Interest

It is compelling to see many interactions of trust and trustworthiness as similar to the interaction between Trifonov and the lieutenant colonel while it was ongoing. The trusted party has incentive to be trustworthy, incentive that is grounded in the value of maintaining the relationship into the future. That is, I trust you because your interest encapsulates mine, which is to say that you have an interest in fulfilling my trust. It is this fact that makes my trust more than merely expectations about your behavior. Any expectations I have are grounded in an understanding (perhaps mistaken) of your interests specifically with respect to me. The relationship of Trifonov and the lieutenant colonel exemplifies a minimal core part of a remarkable array of trust relationships. That minimal core is that there is a clear, fairly well defined interest at stake in the continuation of the relationship.

More generally, it is principally those with whom we have ongoing relationships that we trust. In addition, the richer an ongoing relationship and the more valuable it is to us, the more trusting and trustworthy we are likely to be in that relationship. When asked whom they trust in various ways, people typically name certain relatives, friends, and close associates. The relationship between Trifonov and the lieutenant colonel was a minimal instance of trust in that it was grounded merely in the interest in the ongoing material benefits from the interaction. The lieutenant colonel valued his relationship with Trifonov only for that material benefit. It was specifically his relationship with

Trifonov that mattered to him, however, because this relationship was beneficial to him. His interests and those of Trifonov were causally connected so long as their relationship continued.

While one might object superficially to bringing interests into trusting relationships, such as one's relationship with a close relative or friend, they are clearly there much, and perhaps most, of the time. For many other trusting relationships, the whole point is likely to be interests. For example, I have an ongoing commercial relationship with a local merchant that becomes a trust relationship. For many common trust relationships there is a far richer range of benefits from the relationship than the material interests that motivated Trifonov and the lieutenant colonel, two partners in crime. I enjoy the presence of many people in my life, and I want to maintain my relationships with them. Therefore, they can trust me in various ways. There might even be relationships that I value in themselves and not primarily because they are causally connected to certain other benefits that I get from them. For example, I love certain people and have rich friendships with others. Such relationships are at one extreme of the range of encapsulated interest, and the relationship between Trifonov and the lieutenant colonel is at the other, minimal, extreme. Many of our relationships with others develop from relatively minor exchange interactions and become much richer relationships of fairly broad reciprocity.

Both the relatively limited relationship between Trifonov and the lieutenant colonel and the relatively rich relationship you might have with a friend involve trust as encapsulated interest, which we may characterize as follows: I trust you because I think it is in your interest to attend to my interests in the relevant matter. This is not merely to say that you and I have the same interests. Rather, it is to say that you have an interest in attending to *my* interests because, typically, you want our relationship to continue. At a minimum, you may want our relationship to continue because it is economically beneficial to you, as in the case of Trifonov's relationship with the lieutenant colonel. In richer cases, you may want our relationship to continue and not to be damaged by your failure to fulfill my trust because you value the relationship for many reasons, including nonmaterial reasons. For example, you may enjoy doing various things with me, or you might value my friendship or my love, and your desire to keep my friendship or love will motivate you to be careful of my trust.

Note that *our merely having the same interests with respect to some matter does not meet the condition of trust as encapsulated interest*, although it can often give me reason to expect you to do what I would want you to do or what would serve my interests (because it simultaneously serves yours). The encapsulated-interest account does entail that the truster and the trusted have compatible interests over at least

some matters, but such incentive compatibility, while necessary, is not sufficient for that account, which further requires that the trusted values the continuation of the relationship with the truster and has compatible interests at least in part for this reason. Other drivers on a highway and I enjoy incentive compatibility, and therefore each of us can be expected to try to drive on the appropriate side of the road to avoid accidents with one another. Generally, however, there is no sense in which the other drivers want me to be in the relationship of driving on the same road with them, and therefore I am not in a relationship of trust as encapsulated interest with them.

One could assert a definition of trust that is nothing more than incentive compatibility or rational expectations of the behavior of the trusted. The word trust would be otiose in such a theory, however, because it would add nothing to the somewhat simpler assumption of compatible interests in explaining behavior. The massive literature on trust has not been stimulated by any such simplistic conception of trust. Much of that literature seems to suppose, for example, that there are important normative issues in the seeming fact of declining trust, and much of it supposes that trust is a complex and important matter in its own right. Indeed, Niklas Luhmann (1980, 22, 30) supposes that institutional devices that arrange for merely stable expectations have, of necessity, been substituted for relationships of trust in our complex modern times.

A fully rational analysis of trust would depend not solely on the rational expectations of the truster but also on the *commitments*, not merely the regularity, of the trusted. How can one secure commitments from someone whose love or benevolence does not guarantee good will toward oneself? The most common way is to structure incentives to match the desired commitment. You can more confidently trust me if you know that my own interest will induce me to live up to your expectations. Your trust is your expectation that my interests encapsulate yours. On this view, as Thomas Schelling (1960, 134–35) notes, "trust is often achieved simply by the continuity of the relation between parties and the recognition by each that what he might gain by cheating in a given instance is outweighed by the value of the tradition of trust that makes possible a long sequence of future agreement."

Continuity of the relationship is not enough, of course, because the commitments matter. In a favorite philosopher's example, Immanuel Kant's neighbors may have relied on his punctuality in his morning walk to set their own schedules. To trust him, however, would require more: that they rely on his having their interests at heart in deciding when to take his walk. If they could not think he did, they could not be said to trust him in the strong sense of the encapsulated-interest account (Baier 1986, 234).

Writings on trust often take the view that it involves something beyond merely reasonable expectations based in self-interest.[1] In particular, some writers suppose trust is an inherently normative notion (Elster 1979, 146; Hertzberg 1988). We can make some sense of such claims by supposing they are really misplaced claims about trustworthiness rather than about trust. You might be trustworthy in the strong sense that you would reciprocate even when it was against your interest to do so, as Trifonov might have returned the final 4,500 rubles.

Various social scientific accounts of trust take for granted that trust is rational in the sense of being based on empirically grounded expectations of another person's (or an institution's) behavior (Barber 1983; Luhmann 1980). Trust can lead to intentional or motivational moves by the trusted as well as by the truster. A rational analysis of trust of another intentional being, as opposed to "trust" of a force of nature (our "trust" that it will not rain on a July day in Palo Alto), must take account of the rationality of both intentional parties. Indeed, in a trust relationship, I must think strategically, because my purposes are served by the interaction between what I do and what another does (or others do). My outcome is the joint outcome of both our actions. In this sense, mere expectation accounts are only half strategic, and they therefore fail to address the central nature of trust relationships. They have a liability not unlike that of the similarly half-strategic Cournot theory of market behavior. In the Cournot theory, actors assume regularity of behavior on the part of others in the market in order better to decide how to act themselves; but, although they are strategic in responding to others' actions, they suppose that others are not strategic in responding to them. Hence they fail to take account of second-order effects of others' responses to their actions. Cournot actors are somewhat smart but they think others are dumb.

Many writings on trust convey a vague sense that trust always requires more than rational expectations grounded in the likely interests of the trusted. If this sense is correct, then we are at a very early stage in the development of any theory to account for trust or even to characterize it in many contexts. If an account from interests is largely correct for a large and important fraction of our trusting relationships, however, we already have the elements of a theory of trust that merely wants careful articulation and application. In what follows, I give an account of trust as essentially rational expectations about the self-interested behavior of the trusted. The effort to construct such an account forces attention to varieties of interaction in which trust might arise and hence to differences in the plausible explanations of trust. The sense that trust inherently requires more than reliance on the self-interest of the trusted may depend on particular kinds of in-

teraction that, while interesting and even important, are not always of greatest import in social theory or social life—although some of them are, as is the trust a child can have in a parent.

Elements of Trust as Encapsulated Interest

The encapsulated-interest view of trust includes several elements, some of which are common to other accounts of trust. First, trust is generally a three-part relation that restricts any claim of trust to particular parties and to particular matters. Second, trust is a cognitive notion, in the family of such notions as knowledge, belief, and the kind of judgment that might be called assessment. All of these are cognitive in that they are grounded in some sense of what is true. These cognitive notions—and trust, in particular—are not a matter of choosing: we do not choose what is to count as true, rather we discover it or are somehow convinced of it. Hence we do not trust in order to accomplish anything, although our trust might encourage us to enter beneficial interactions. (We may well choose to be trustworthy in order to encourage others to cooperate with us.) Generally, we wish to explain cooperation or its failure by reference to trust. To make the cooperation itself a matter of trust would make the thing we want to explain the explanation of it. Thus we wish to keep trusting and acting from trust cleanly separated. Finally, acting on trust typically involves risk.

Other issues of trust need be only mentioned, not discussed at length here. First, trust involves expectations of behavior from another, but not just any expectations. The expectations must be grounded in the trusted's concern with the truster's interests. Second, trust and trustworthiness are subject to the larger context. Your encapsulation of my interest in making your own choices may not be sufficient to get you to fulfill my trust because other considerations may trump. For example, two people might trust you with respect to different things, and in fulfilling one of those trusts you might violate the other. Taken out of context, your trustworthiness in each of these relationships might be in your interest. But when they come into conflict in the context of your wider life, one interest might trump the other.

Many writers take issue with one or another of these elements of trust. Some of the criticisms appear to be matters that are normative, as in some of the views canvassed in chapter 3, where I take up alternative conceptions of trust. Some of the disagreements, however, are genuinely conceptual. The view for which I argue here seems to fit modal cases of actual trusting, in which the trusting makes a real difference to how people then behave. It also seems to yield or fit

with an explanation for both the trusting and the behavior that follows from it.

One other element of trust is shared in virtually all views: competence to do what one is trusted to do. You should not trust me to get you safely to the top of Mount Everest and back, even if I convince you that I have the best will in the world to do so. I usually assume throughout this book that competence is not at issue in the trust relationships under discussion. The point of this is not to dismiss the problems of competence and of judging someone's competence—such problems are often severe and de facto insurmountable barriers to trust—but merely to concentrate on motivational issues. There are, of course, many contexts in which competence is a major issue as well as many in which the problem of knowing how competent someone is can be very difficult. A substantial book could be written on these issues, but this is not that book.

Competence is a major issue in many contexts in which specialized abilities are at issue, as is typically true of professional services as well as ordinary individual interactions. You would probably have less confidence in the competence of a young and inexperienced teenager as a baby-sitter than in that of an older and experienced person. In such a case, you might know enough to judge the relative competence of these two people. In other contexts, however, the issue is how to judge someone's competence. We commonly prefer to call on people whom we know to be competent and avoid relying on those about whom we know too little to judge them.

My competence in getting up Mount Everest is, of course, a fairly fixed characteristic that is not specifically mobilized to answer to your potential trust in me. If I have not already developed such a capacity, you should not want to rely on my somehow developing it while leading you up that mountain. Most of the motivational issues in trust—for the encapsulated-interest account as well as for most others—are much more clearly specific to the particular relationship at issue. Again, therefore, the focus of this book is on motivations, which are far less well understood in the trust literature than is the problem of competence.

Certain institutional arrangements convert our particular personal judgment problems into problems of generalized assessments. For example, we have agencies that assess the competence of such professionals as doctors, lawyers, and even mountain-climbing guides. These agencies also commonly oversee motivational commitments of the professionals—for example, they attempt to regulate conflicts of interests. Testing and certification of competence is, however, a major part of their task. Such agencies convert our relations with professionals into something different from the kind of trust relationship I

might have with you personally. Indeed, they arguably eliminate much of the trust we might otherwise have developed, so that our dealings with professionals have more the character of assessing and acting on mere expectations. Similar devices of third-party "certification"—as by a mutual friend or a Chinese guanxi mediator—also often stand in for direct assessments of those we must rely on for ordinary personal relations.

Trust as a Three-Part Relation

A characteristic of trusting relationships, one that is not uniquely relevant to the encapsulated-interest view, is that trust is generally a three-part relation: A trusts B to do X (Baier 1986; Luhmann 1980, 27).[2] Even then, the trust depends on the context. For example, I might ordinarily trust you with even the most damaging gossip but not with the price of today's lunch (you always—conveniently?—forget such debts), while I would trust another with the price of lunch but not with any gossip. I might trust you with respect to X but not with respect to ten times X. Some few people I might trust with almost anything, many others with almost nothing. But in a radically different context, such as when you are under great duress and my piece of gossip would help you out of a bad situation, I might no longer trust you with it.

To say "I trust you" seems almost always to be elliptical, as though we can assume some such phrase as "to do X" or "in matters Y."[3] Only a small child, a lover, Abraham speaking to his god, or a rabid follower of a charismatic leader might be able to say "I trust you" without implicit modifier. Even in their cases we are apt to think they mistake both themselves and the objects of their trust. Many of us, of course, might start by taking a risk on newly encountered people or people in newly undertaken areas, but we would prefer not to take such a risk in important matters without a substantial prior history of trustworthiness and a strong sense that the trusted will have incentive to follow through.

Those who see trust as normative or otherwise extrarational argue that it is more richly a two-part or even one-part relation than this view implies. It is a one-part relation if I trust out of a pure disposition to trust anyone and everyone with respect to anything and everything, in which case I am the only variable part. There may be people, especially children, naïve enough to have such a disposition, but most of us clearly do not have it. There is a fairly extensive literature on so-called generalized trust, which is trust in the general other person whom we might encounter, perhaps with some restrictions on what matters would come under that trust. Conceptual issues in sur-

vey research on generalized or social trust are discussed in chapter 3 (also see the appendix) and the implications of the results of such research in chapters 7 (trust in government) and 8 (trust and society). But here note that this category has two odd features. First, it sounds more nearly like a simple expectations account than a richer trust account. In this account, I supposedly think everyone is reliable up to some degree independently of who they are or what relationship I have with them. I think this of them the way I might think the typical person would behave in certain ways in various contexts.

Second, when survey respondents say they trust most people most of the time, this is almost surely an elliptical claim. They do not mean that, if a random stranger on the street were to ask for a loan of, say, a hundred dollars, they would trust that person to repay and would therefore make the loan. This ellipsis might be covered by the phrase "most of the time." Hence even this open-ended answer to a badly framed, vague question is almost certainly just a loose way of saying they would trust most people within somewhat narrow limits. Moreover, it is also elliptical in its reference to "most people." Few of the respondents would genuinely trust just anyone much at all.

Trust and Cooperation

Trust is in the cognitive category with knowledge and belief. To say I trust you in some way is to say nothing more than that I know or believe certain things about you—generally things about your incentives or other reasons to live up to my trust, to be trustworthy to me. My assessment of your trustworthiness in a particular context is simply my trust of you. The declarations "I believe you are trustworthy" and "I trust you" are equivalent. If it is cognitive, it follows that trust is not purposive (Baier 1986, 235). I do not trust you in order to gain from interacting with you. Rather, because I do trust you, I can expect to gain from interacting with you if a relevant opportunity arises. Moreover, if trust is cognitive, it is not behavioral. I may act from my trust, and my action may give evidence of my trust, but my action is not itself the trust, although it may be compelling evidence of my trust.[4] If I trust you, I trust you right now and not only in some moment in which I act on my trust by taking a risk on you.[5]

Suppose my trusting you in some matter is rational in the sense of being well grounded. What follows? Perhaps nothing. That I trust you does not entail that I should act on the trust. There might be other things I would rather do at the moment or other people whom I similarly trust for the matter at hand. Therefore, I face a choice of what to do even though it is incoherent to say I choose to trust you. If I trust you, I will think it not very risky to rely on you in some matter. (I return to this issue in chapter 3.)

I can, however, also choose to take the risk of cooperating with you on some matter even if I do not trust you. While he was the prime minister of Israel, Ehud Barak, when asked if he trusted Yassir Arafat, said, "I don't know what it means to trust. He is the Palestinian leader, not the Israeli leader, and he is determined to do whatever he can to achieve Palestinian objectives. The real question is not whether we trust him. The question is whether there is a potential agreement that could be better overall for both sides, a win-win, not a zero-sum game" (quoted in Goldberg 2001, 66). Cooperation or coordination is the general goal, but there are many ways to achieve it, some of which do not depend on trust. Hence my actions are not simply determined by the degree of my trust, although they are often likely to be influenced by my trust or distrust.

In the encapsulated-interest account of trust, the knowledge that makes my beliefs about you a matter of trust rather than of mere expectations is my beliefs about your incentives toward me in particular. These are not merely bald, unarticulated expectations about your behavior. I have bald expectations that the sun will rise tomorrow, and I might not be able to give any account of why I think that, other than induction from the past. (As a physicist, you might be able to give a very good account, so that your expectation of the sun's rising is theoretically grounded.) What matters for trust is not merely my expectation that you will act in certain ways but also my belief that you have the relevant motivations to act in those ways, that you deliberately take my interests into account because they are mine.

It is common in the vernacular to say I "trust" you to do such things as, for example, defend yourself if attacked by a dog, in which case your motivations are not at all like those of the encapsulated-interest account of trust. You defend yourself, as most of us would, for your own direct interest. If trust reduces to such bald expectations of behavior, there is little point in using the loaded term "trust." My "trust" would be useless in helping us explain your self-defense, which is not motivated by your concern with my interests (or any other commitments you might have to me specifically). Moreover, my trust—as my assessment of your encapsulation of my interests in your own interests—will commonly help explain relevant actions of mine, specifically my choosing to rely on you to do something on my behalf, whereas my "trust" that you would defend yourself when attacked by a dog would explain none of my behavior.

Acting on Trust as Involving Risk

As virtually all writers on trust agree, acting on a trust involves giving discretion to another to affect one's interests. This move is inherently subject to the risk that the other will abuse the power of discre-

tion. As David Hume (1978 [1739–40], 3.2.2: 497) observes, "'Tis impossible to separate the chance of good from the risk of ill." Hence to act on trust is to take a risk, although trust is not itself a matter of deliberately taking a risk because it is not a matter of making a choice.

As an objection to the encapsulated-interest account, one might suppose it perverse to say I trust you to do X when it is in your interest to do X. For example, consider an extreme case: I am confident that you will do what I want only because a gun is pointed at your head. (I have grasped the wisdom of Al Capone, who is supposed to have said, "You can get so much farther with a kind word and a gun than with a kind word alone" [McKean 1975, 42n]).

My coercing you to do what I "trust" you to do violates the sense that trust has no meaning in a fully deterministic setting. I do not trust the sun to rise each day, at least not in any meaningful sense beyond merely having great confidence that it will do so. Similarly, I would not, in our usual sense, trust a fully programmed automaton, even if it were programmed to discover and attempt to serve my interests—although I might come to rely heavily on it. Many writers therefore suppose that trust is inherently embedded in uncertainty. "For trust to be relevant," Diego Gambetta (1988, 218–19) says, "there must be the possibility of exit, betrayal, defection" by the trusted (see also Yamagishi and Yamagishi 1994, 133; Luhmann 1980, 24). More generally, one might say trust is embedded in the capacity or even need for choice on the part of the trusted. Giving people very strong incentives seems to move them toward being deterministic actors with respect to the matters at stake. At the other extreme, leaving them with no imputable reasons for action generally makes it impossible to trust them. Trust and trustworthiness (and choice and rationality) are at issue just because we are in the murky in-between land that is neither deterministic nor fully indeterminate. Yet it still can make sense to say of someone, such as your mother, that you trust her virtually beyond doubt with respect to very many things. Such people, however, are rare in our lives. Trust is a problem of often great interest just because so few of our relationships are like that one.

Part of the issue in the gun case is that your compliance with my request is not motivated by your concern with my interest at all. It is motivated purely by your concern with your own interests. Hence the gun case fails to fit the encapsulated-interest account of trust, which would require your concern with my interests. Luhmann (1980, 42; see also Hertzberg 1988) seemingly opposes the encapsulated-interest account because it turns on the interests of the trusted. "It must not be that the trusted will toe the line on her own account, in the light of her interests," he writes. This unexplicated obiter dictum runs counter to his own general account, according to which the overriding consid-

eration is that the two parties in a trust relation are typically going to meet again (Luhmann 1980, 37)—presumably in an iterated or ongoing exchange relationship in which a strong reason for trustworthiness is one's interest in keeping the relationship and its exchanges going. His claim, however, might be the misstated observation that trust must not be a matter of the trusted's acting only on his or her own account without reference to the interests of the truster.

The Rationality of Trust

At the individual level, my trust of you must be grounded in expectations that are particular to you, not merely in generalized expectations. If I always trust everyone or if I always act from generalized expectations, then I do not meaningfully trust anyone. Trust is therefore in part inherently a rational assessment. My expectations about your behavior may be grounded in my belief in your morality or reciprocity or self-interest. With no prior knowledge of you, I may initially risk treating you as though I trust you, but our relationship can eventually be one of trust only if there are expectations that ground the trust. As Karamazov's lieutenant colonel learned, expectations that are well grounded in one context may not be reliable for new contexts, such as his sudden loss of status as base commander.

That trust is essentially rational is a common view. For example, James Coleman (1990, chapter 5; also see several contributions to Gambetta 1988) bases his account of trust on complex rational expectations. There are two central elements in applying a rational-choice account of trust: incentives of the trusted to fulfill the trust and knowledge to allow the truster to trust. The knowledge at issue, of course, is that of the potential truster, not that of the theorist or social scientist who observes or analyzes trust. Hence we require an account of the epistemology of individual knowledge or belief, of street-level epistemology, to complete the rational theory of trust (see chapter 5).

A full statement of the rational theory, including the incentive and knowledge effects, is as stated earlier: Your trust turns not directly on your own interests but rather on whether these are encapsulated in the interests of the trusted. You trust someone if you believe it will be in her interest to be trustworthy in the relevant way at the relevant time, and it will be in her interest because she wishes to maintain her relationship with you. Some accounts of trust do not specifically include reference to the trusted's interest in being trustworthy toward the truster but merely require an expectation that the trusted will fulfill the trust (Barber 1983; Gambetta 1988, 217–18; Dasgupta 1988). Adequate reason for such an expectation, however, will typically turn on an assessment of likely future incentives.[6]

The encapsulated-interest account backs up a step from a simpler expectations account to inquire into the reasons for the relevant expectations—in particular, the interests of the trusted in fulfilling the trust. The typical reason for the expectations is that the relations are ongoing in some important sense. There are two especially important contexts for trust: ongoing dyadic relationships and ongoing—or thick—group or societal relationships. The two classes are closely related, and both are subsumed in the encapsulated-interest account of trust. The first class is divisible into one-way and mutual trust relations, both of which are grounded in ongoing dyadic interactions. Such interactions pose incentives to the trusted that are of increasing severity. The sanction that compels the trusted party in a one-way trust game and both of the trusted parties in the mutual trust exchange interaction is withdrawal by the other party and therefore the loss of future benefits from the interaction. The sanction in thick relationships can go beyond such withdrawal to include shunning from the whole community of those who share in the thick relationships. Let us consider each of these in turn.

One-Way Trust

The interaction of Trifonov and the lieutenant colonel was, the first time they dealt with each other, an instance of what we may call the one-way trust game. This standard game has been widely used for the experimental study of trust for about a decade (Kreps 1990; McCabe and Smith in press; Hardin in press a). (Variants of this game with other payoffs are strategically identical in the sense that the orders of the payoffs are the same.) The lieutenant colonel must act as though he trusts Trifonov in order to gain from their interaction, whereas Trifonov need only act in his own interest. The game illustrates one-way trust because it is only the lieutenant colonel whose actions might depend on his trusting. The lieutenant colonel can never cheat Trifonov. In the game, the lieutenant colonel makes the first move of lending or not lending the rubles. If he does not lend, the game ends with payoffs of nothing to both parties. If he lends, then there follows a next stage in which Trifonov chooses whether to repay fully with an additional personal gift or not to repay. The play of the game ends with his choice.

If the game is played only once, clearly Trifonov's interest is to defect, and therefore the lieutenant colonel's interest is not to make the initial loan. In many experiments using variants of this game, the first mover (the lieutenant colonel here) often risks cooperation and the second mover often reciprocates, even when the game is to be played only once (which appears to be the most common way to use

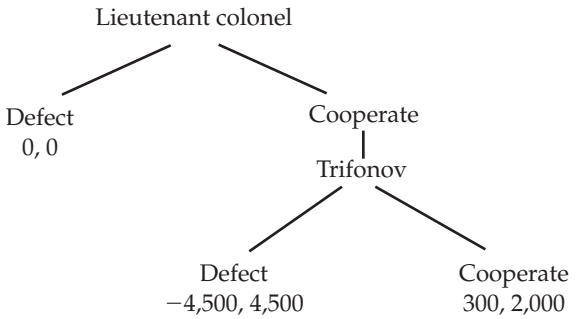
the game experimentally). If the experiments were run with payoffs on the scale of what the lieutenant colonel and Trifonov faced, we might expect almost no cooperative plays in games played only once. If, however, the players were like the lieutenant colonel and Trifonov and were able to play the game repeatedly over the years, their interests might change dramatically, because both could do very well over many plays of the game.

Let us suppose that after every audit, there will be about 4,500 rubles to play with and that Trifonov makes a generous profit from his investments, enough to yield 2,000 rubles to himself and a gift of 300 rubles to the lieutenant colonel. The payoff structure of a single play of their game will then be as in game 1 (figure 1.1). Moves are sequential. First, the lieutenant colonel must choose either to defect or to cooperate. If he chooses to cooperate, then Trifonov must decide whether he will defect or cooperate. If Trifonov gets the loan three times and successfully invests it and repays it, he makes a clear profit (1,500 rubles) in comparison with cheating already on the first loan of 4,500 rubles. If he carries it off often for several years, he makes a very large profit that swamps the initial 4,500 ruble loan. Repayment, with the small gift, is therefore clearly in Trifonov's long-run interest. Presumably Trifonov can cheat the lieutenant colonel only once, and while the lieutenant colonel is powerful, Trifonov might suffer reprisals if he cheats. Hence so long as he can foresee two more plays of the game beyond the current play, it always serves his interest to repay the money. Once the relationship is clearly over, however, and there is no longer any chance of reprisal, there is only a short-run gain from a single final loan, and that is substantially trumped by the gain from cheating and keeping the 4,500 rubles (plus any profit he has made from them).

The game as represented might not include all the relevant payoffs. For example, so long as he has the resources of his power over the military base and his standing in the local community, the lieutenant colonel has power to take vengeance on Trifonov if Trifonov cheats. When he was embarrassed by an unannounced audit and discovered that he would be replaced in office, however, he lost that power. If the initial loan had been a personal loan from the lieutenant colonel's own funds, it could have been governed by a legally enforceable contract, so that trust need have played little or no role in the lieutenant colonel's expectation of getting full repayment with a bit of interest.

Note, incidentally, that the one-way trust game of game 1 is a three-part relation. The lieutenant colonel trusts Trifonov with respect to about 4,500 rubles. This is true for the general trust game no matter who the parties are or what the stakes are. In addition, it would, of course, be plausible for the lieutenant colonel not to trust Trifonov in

Figure 1.1 Game 1: One-Way Trust



Source: Author's configuration.

a one-way trust game with radically higher stakes even though he did trust him with the stakes at 4,500 rubles.

It is a great strength of the experimental protocol for the trust game that it virtually forces us to be clear on at least some of what is at issue. It is difficult to imagine a reduced analog of the one-way trust game that would represent only a two-part relation unless it allowed the payoffs to be merely ordinal and completely open-ended. In that case, however, the relevant player would be unable to choose to cooperate at the first move because the loss, if the other party chose to take the noncooperative payoff, could be catastrophic. Players who understood such a game could not, if their own resources were at stake, seriously claim to think it smart to cooperate at the first move. Unlike the findings of experiments using these games, survey results on so-called generalized trust can be based on questions that are vague and even glib and can therefore confuse what is at issue (see the discussion in chapter 3 and the typical survey questions presented in the appendix).

As the example of Trifonov and the lieutenant colonel shows, even when it is played once only, the one-way trust game represents real choice problems. The choices precede any trust relation, however; hence calling it a trust game is misleading if the game is not iterated. The example is also in a sense only half of commonplace trust relations, in which both parties are at risk, both might trust or not trust, and both might be trustworthy or not trustworthy. For example, in an ongoing mutual exchange relationship, you and I might both be in a position on occasion to cheat each other. Neither of us would have the restricted role of the lieutenant colonel, who can trust or not trust but cannot act on the misplaced trust of Trifonov, because Trifonov need

not trust. Let us turn to this slightly more complex case of mutual trust, in which the two parties are in a symmetric relationship.

Mutual Trust

Iterated one-way trust relationships are of great analytical interest because of their simplicity, although they are arguably somewhat unusual in the mass of all trust relationships. One can imagine that some parent-child relationships are virtually one-way relationships, as are many others in which the parties are not equal or are not in symmetric roles. For good reason, the more stable and compelling trust relationships are likely to be mutual and ongoing. Why is this so? Because a good way to get me to be trustworthy in my dealings with you, when you risk acting on your trust of me, is to make me reciprocally depend on your trustworthiness. A reciprocal trusting relationship is mutually reinforcing for each truster, because each person then has built-in incentive to be trustworthy (Coleman 1990, 77). I trust you because it is in your interest to do what I trust you to do, and you trust me for the reciprocal reason. If, as subjectively seems to be true, trust relationships are typically reciprocal, we have reason to suppose they are not typically grounded in particular characteristics of the trusted. They are relational because they are grounded in incentives for trustworthiness, as in the encapsulated-interest account.

The prototypical case of mutual trust at the individual level involves an interaction that is part of a long sequence of exchanges between the same parties. Each exchange is simply the resolution of a prisoner's dilemma (Hardin 1982b). A sequence of exchanges is therefore an iterated prisoner's dilemma with, perhaps, some variation in the stakes at each exchange. Hence the main incentive that one faces in a particular exchange in which one is trusted by the other is the potential benefit from continuing the series of interactions. The sanction each of us has against the other is to withdraw from further interaction.

The model of mutual trust as trust from iterated exchange is not a definition of trust in certain relationships, but it is an explanation of much of the trust we experience or see, much of which is reciprocal and is grounded in ongoing relationships. As discussed further in chapter 3, trust is typically reducible to other terms. We can determine what some of these other terms are from the iterated exchange model, which is an explanatory theory of trust. Ongoing dyadic relationships of trust typically involve mutual trust.[7]

Ordinary exchange can be represented as a prisoner's dilemma game, as in figure 1.2 (see Hardin 1982b). In this game, the payoffs to each player are strictly ordinal. They bear no relation to dollars or

Figure 1.2 Game 2: Prisoner's Dilemma or Exchange

		Column Player	
		Cooperates	Does not cooperate
Row Player	Cooperates	2, 2	4, 1
	Does not cooperate	1, 4	3, 3

Source: Author's configuration.
(x,y)
x = row player
y = column player
Note: In each cell, the first payoff is the row player's, the second the column player's.

utiles. They merely indicate the order of optimal benefit for each player from each possible interaction. The first cell, for example, indicates that the cooperation of both parties yields the second-best payoff for each of the players. The outcome with a payoff of 1 is the player's first choice, or most preferred outcome, that with a payoff of 2 is the player's second choice, and so forth. There is therefore no sense in which we can add, say, the payoffs that are ranked 1 and 4; nor can we say that Row's 1 is comparable in magnitude to Column's 1. In each cell of the matrix, the first payoff goes to the row player and the second to the column player (in the mnemonic Roman Catholic convention). Hence the top left cell of the game gives both players their second-best outcomes, which are an improvement over their status quo third-best outcomes that result from joint failure to cooperate with each other. (Such games are more commonly presented with cardinal payoffs in money rather than with merely ordinally ranked outcomes.)[8]

If we play the prisoner's dilemma once only with no expectation of encountering each other again in an exchange relation and without the benefit of any external agency to compel us to cooperate, it is in our interest individually not to cooperate. If we play the game repeatedly, however, we have strong incentive to cooperate, if we can get each other to recognize this fact. Therefore, the once-only interaction has none of the force of the encapsulated-interest account to get us to trust each other, but an iterated, ongoing interaction does have that force (Hardin 1982a, chapters 9 to 14). Some game theorists argue that iteration cannot generate incentives to cooperate in ongoing interactions in the ordinary prisoner's dilemma. I briefly address their objection later in this chapter.

Several types of behavior often identified as moral can be clearly understood as self-interested in many contexts. Promise keeping, honesty, and fidelity to others often make sense without any presupposition of a distinctively moral commitment beyond interest. Consider promise keeping, which has been the subject of hundreds of articles and books in moral theory during the past century.[9] In the eighteenth century, David Hume (1978 [1739–40], 3.2.5: 523) said, without seeming to think the statement required much defense, that the first obligation to keep a promise is interest. The claim is obviously true for typical promises between close associates who have an ongoing relationship that they want to maintain. If I promise to return your book, I will be encouraged to do so by frequent contact with you and frequent desire to make other exchanges with you. If I generally fail to keep such promises, I can probably expect not to enjoy as many exchanges and reciprocal favors. Promising relationships typically are those in which exchanges are reciprocated over time. Because exchanges are resolutions of prisoner's dilemma problems, promising relationships involving exchange have the incentive structure of iterated plays of the prisoner's dilemma.[10] Prima facie, it is in one's interest to keep such a promise, although that interest might be trumped by some other (see further Hardin 1988b, 41–44, 59–65). (I discuss the relationship between promise keeping and trust in chapter 3.)

A strong external force generally backs promises: the loss of credibility that follows from breaking them. Without credibility, one loses the possibility of making promises. Why should anyone want the power to make promises? All I really want in my own interest is the power to receive them. And there's the rub, because promises are generally part of a reciprocal exchange. The real penalty here is not that others will no longer rely on me but that they will not let me rely on them. As is commonly true also of trust relationships, promising typically involves intentions on the parts of two people. As with promising, future expectations, generally based in ongoing experience, contribute much of the force that binds in a trusting relationship. Trifonov and the lieutenant colonel could trust each other so long as future expectations of their relationship were motivating.

When it is repeated, the one-way trust game has some of the quality of the iterated prisoner's dilemma and therefore of mutual trust. It is not a prisoner's dilemma, however, *because there is no outcome that is best for the lieutenant colonel that is simultaneously worst for Trifonov*. When played once, the prisoner's dilemma of game 2 (figure 1.2) has four outcomes, whose orderings of payoffs define the prisoner's dilemma, while the one-way trust game has only three outcomes. The worst outcome for Trifonov in any given play of the one-way trust game is analogous to the noncooperation, or status quo, outcome (3,

3) in the prisoner's dilemma. If Trifonov is the column player in game 2, the outcome in which the lieutenant colonel's payoff is 1 and Trifonov's 4 is not possible; hence a two-by-two matrix representation of the one-way trust game has an empty cell. Still, there must be some degree of mutual trust if they are to continue playing because each time he returns the 4,500 rubles Trifonov must take a risk that the lieutenant colonel will not continue the arrangement after the next and other future audits.

In mutual trust, again, the interaction is a finitely iterated exchange or prisoner's dilemma. According to a standard argument in game theory, one should not cooperate in such a game. The argument begins with the premise that one should treat the final play of a finite series of plays of the game as a one-shot game, in which one should defect. If one should defect on the final play, however, then the penultimate play is de facto a final play in the sense that it can have no effect on anything thereafter, and so one should defect on the penultimate play as well. By tedious induction backward, one should defect already on the first play in the series.

If the backward induction argument is compelling, it is hard to see how rational individuals could ever enter into normal relationships of trust and exchange. All such relationships would have to be grounded in something extrarational, perhaps in normative commitments to be altruistic or more decent than is rational. On this view, the fact that there is apparently a great deal of trust in our lives suggests that we are not rational. I think, on the contrary, that trust is eminently rational and that the backward induction argument is flawed. In brief, the flaw is this: Suppose I know that you are eminently rational and that you believe the backward induction argument. I also know that we could gain substantially from entering a series of exchanges that must terminate, perhaps unhappily, at some distant future point. I can now wreck your backward induction by simply cooperating at our first encounter. You may now suppose I am irrational, or you may reconsider your induction. Either way, you may now decide it is in your interest to reciprocate my cooperation, so that we both gain far more than we would have from continuous mutual defection. Indeed, I think you must reconsider your induction because if I, by acting cooperatively, can get you to cooperate, you should realize that you could do as well with others in such an interaction. That is to say, you must agree that it would be sensible for you to cooperate initially rather than to defect.[11]

Moreover, and more to the point here, if you think cooperation in finitely iterated prisoner's dilemma interactions is irrational, you must wonder at your own tendency initially to take risks of cooperating with those whom you do not yet know well. All our relationships

with people are of perhaps ill-defined but necessarily finite duration. The backward induction argument recommends initial distrust and, furthermore, continued distrust. This is a recommendation for slow death by abnegation in mimicry of Herman Melville's (1984) Bartleby, the scrivener, who became so asocial that he died of starvation and whose response to every entreaty was, "I would prefer not to." Whatever the apparent force of the backward induction argument for rarefied game theorists, it appears that actual people in living societies, including the game theorists who preach against the rationality of doing so, regularly take the risk of initially cooperating to upset that argument. Only for that reason do we have living societies.

The analysis here of the iterated prisoner's dilemma applies as well to an iterated one-way trust game, such as that between the lieutenant colonel and Trifonov. When that game is iterated, Trifonov has reason to cooperate in order to induce the lieutenant colonel to continue to loan him the loose cash after each periodic audit.

Thick Relationships

Now turn to trust that is grounded in a complex of overlapping iterated interactions over broad ranges of matters. In a small, close community, each of us can have ongoing relationships with every other one of us. Such overlapping relationships typically generate a lot of knowledge relevant to trusting any particular person, and they generate incentives not only between two partners in trust but also between each of them and others in the thick community. Even outside such a close community, I may belong to a subcommunity of similarly overlapping relationships with a close circle of relatives and friends and a small number of others with whom I regularly deal. In our subcommunity we may all know one another well enough to know the limits of one another's trustworthiness and to rely on each member's being responsible not merely to a particular truster but to the entire group of us. Those with whom we deal have not only the incentive of loss of our relationship but also that of loss of reputation and the possibility of shunning by others if they cheat us on a deal. Among these people we therefore know whom we can trust for what (Williams 1988). We may say that trust in these contexts of a close community or subcommunity builds on thick relationships.

Bernard Williams (1988) explicitly and some others implicitly define trust as a function of thick relationships. Williams supposes that therefore trust is not possible in many contexts in which we do not have such relationships. For example, he views the issue of my trusting political leaders as though it were an exact analog of the more familiar problem of my trusting a close associate. Because I am not

involved in many overlapping interactions with the typical political leader, then, on the thick-relationship theory I cannot trust him or her; hence trust cannot handle this relationship in general (Luhmann 1980).

Williams, Luhmann at times, perhaps the anthropologist F. G. Bailey (1988), and others seem to see thick relationships as virtually definitive of trust. The correct way to see the role of thick relationships, however, is as one possible source of knowledge for the truster about the trustworthiness of another and one possible source of incentives to the trusted to be trustworthy. The first of these is essentially an epistemological role. Obviously, however, thick relationships yield only a part of the knowledge we have of others. Our understanding should not stop with only the thick-relationship class of epistemological considerations. In practice, this class may often have priority among our sources in our face-to-face interactions, but this descriptive fact does not give it conceptual or theoretical priority. A fully articulated theory will include this class as a part, not as the whole story, of the epistemology of trust. There is unlikely to be any quarrel with the view that knowledge of another's trustworthiness can come from many sources other than thick relationships.

Similarly, a thick relationship with another is only one of many possible ways to give that other the incentive to be trustworthy. A thick relationship with the truster commonly gives the trusted such incentives not only through the workings of an iterated prisoner's dilemma of reciprocal cooperation but also through reputational effects on others in the thick community. Such reputational effects must have a substantial effect on trustworthiness among familiar relations. Reputational effects give me an incentive to take your interests into account even if I do not value my relationship with you merely in its own right. They do this indirectly because I value relationships with others who might react negatively to my violation of your trust. Because my reputation is valuable to me in my further relationships, I encapsulate your interests in my own to some extent. The thick-relationship theory is therefore merely a special case of the encapsulated-interest theory of trust. It is a partial theory that does not generalize to many contexts. In any theory of trust, the restriction to small-scale thick relationships must follow from other principles. Going back to those principles is a first step in generalizing the theory.

It is a merit of the thick-relationship theory of trust that it blocks the quick blurring of individual and institutional problems, which is one of the most common mistakes in writings on trust. Writers in all disciplines occasionally succumb to the easy analogy from individual to institutional issues that abstracts from the differences in individual-

level and institutional-level constraints and possibilities. For some explanatory theories of trust and how it can work, Williams's conclusion that trust cannot be generalized beyond the small scale may well follow. For other theories, it might be easy to see how individual-level and institutional-level trust are conceptually related even though different kinds of data or evidence are commonly relevant at different levels. I address this issue further in chapter 7.

From Interests to Well-Being

Framing an account of trust as encapsulated interest may provoke an unfortunate misunderstanding. Sometimes interests are the whole story of a person's motivations in a particular context. Typically, however, I have an interest in having more resources, such as money, only because they enable me to consume or experience various things. These consumptions constitute my welfare. The whole story is one of well-being through the use of resources. Interests are merely a proxy for this whole story. It would be a mistake, however, to suppose that interests translate smoothly into well-being or even consumptions. Consumptions generally trade off against one another (and against interests), because if I use my resources for one consumption I may have none for other consumptions.

It is also a mistake to suppose that my well-being is merely selfish. Among the things that make me enjoy life are the enjoyments of certain others. I might enjoy a lovely dinner, but I might enjoy it even more with you. Or I might want my son to enjoy the evening and might use some of my resources to make that possible. My well-being will often depend on my sharing intentions with you to do things with or for you.

It is common to say that people are rational in some contexts and not in others. One might be seemingly rational in choosing between two jobs but not in choosing a spouse. It is even supposed that some whole cultures are less rational than others. James Scott (1976) has argued that the peasants of Southeast Asia, for example, are driven by a "moral economy." What he means is that they do not maximize their production of rice. Rather than adopting seed grains that would have large average annual yields, they stick with grains that will almost always produce enough to keep them from starving but much less on average than the most productive seeds.

Scott says that these peasants have "preferences which do not make sense in terms of income alone" (Scott 1976, 35). But preferences make sense only over whole states of affairs, in which income is only part of what matters. The peasants are like anyone else; they want income only for what it will buy for them. If in a bad year with the

higher-yielding grain they starve, their income will have done little for them. As is presumably true of Scott and virtually everyone else as well, I also do not have an unrestricted preference for higher income. If higher income entails giving up my academic life or more of my leisure time, I may not prefer it to my present income with my present lifestyle and consumption pattern. There are *no preferences* that "make sense in terms of income alone." Income is just a proxy for what we really want. And interest in the encapsulated-interest model of my trust of you is merely a proxy for all that you might take into account on my behalf.[12]

Concluding Remarks

The encapsulated-interest account of trust holds that the trusted encapsulates the interest of the truster and therefore has incentive to be trustworthy in fulfilling the truster's trust. The encapsulation happens through causal interactions in the iterated one-way trust game, iterated exchange (or prisoner's dilemma), and thick relationships. None of these, however, is itself definitive of the trust relation. They are all merely ways to give the trusted incentive to take the interests of the truster into account. This might be done in other ways as well. For example, we might suppose that a near variant of the iterated exchange relationship is reputational effects on my incentives (as discussed further in chapter 6). If I fulfill your trust, that action might help me in other relationships that I value or would value, and if I fail your trust, that action might jeopardize other relationships I might have.

Consider another class of ways I might come to take your interest into account. If I love you, or am your close friend, or am altruistic toward you, I might directly count your interest to some extent as my own. In economists' jargon, I might partially include your utility in mine. Hence you can trust me to some extent just because the effect of our interaction on your welfare will matter to me. We commonly trust our parents, siblings, close friends, spouses, and others who are close to us in this way within varying limits. One might wish to call these normative instances of trust. But the actual trusting is not different from the purely interested cases under the trust game or iterated exchange. If there is a normative quality to these instances from love and so forth, it is in the fact of the love or friendship and the caring for another that follows from these.

It might also happen that what affects your welfare similarly affects mine, so that I should act de facto in your interest just because the same action would be in my interest. Here, however, we would not want to count me as someone you could trust so much as merely

someone from whom you can expect beneficial actions. For example, in the important coordination of traffic, as noted earlier, with everyone driving on the right (or everyone on the left), we share an interest to such an extent that our welfares are causally interdependent even though we need not care at all about one another. My driving on the right is not an instance of my having a positive causal interest in your actions in the sense that I actually want you to interact with me, as I do want you to interact with me in a beneficial exchange and as the lieutenant colonel wanted Trifonov to interact with him. It is, rather, a case in which I would actively prefer that you and I were not even interacting—I would be safer if you were off the road. If your interest is to do what you do independently of my presence, your interest does not meaningfully encapsulate mine.

If we consider all the trust relations we experience, we find that a large fraction of them fall into three categories: relationships or interactions that are iterated, those that are backed up by institutions, and those that are mediated by other (noninstitutional) third parties. This chapter has focused on the first of these categories. Chapter 4, on distrust, suggests that such interactions are inadequate to secure cooperative behavior in many contexts. Chapter 6 considers mediation by third parties (often institutional third parties), and chapters 7 and 8 consider social and institutional devices for mobilizing cooperation where trust might be lacking or inadequate to secure cooperation. All of these categories can be understood easily without any supposed residue beyond rational expectations grounded in the motivations or interests of the cooperating parties because each of them builds in the incentives necessary to induce trustworthy behavior—although, of course, these incentives can be and sometimes are trumped by others. The first category—iterated interactions—seems to be far and away the largest in ordinary interpersonal life. That is because much of our lives is spent in ongoing relationships, such relationships constitute much of what is most valuable to us, and we make substantial commitments to one another in such relationships.

Some of the alternative visions of trust (canvassed in chapter 3) are plausible accounts of some instances of trust. I might trust you with respect to certain things because your moral commitments make you reliable or because your character virtually ensures your relevant action. In such cases, my trust is grounded in an account of your trustworthiness. Such accounts differ significantly from the model of trust as encapsulated interest, and they often have different implications in various social contexts. Consider two examples.

I once had an acquaintance of whom many people said, with genuine force, that he was a person you could trust. Alas, that depended on who "you" were. Many people did not trust him at all because

they thought him deceitful and manipulative. The latter group included people whose interests often conflicted with his and whose future value to him he had seemingly written off. He could be richly and deeply trusted by those who shared enough of his interests, not at all by those who did not. He was almost mythical in his capacity to put people into two distinct classes. On a cynical reading, he was not trustworthy on the encapsulated-interest account. He was merely reliable to those whose interests he happened to share.

For the second example, one might note that a member of some ethnic or other group is extremely reliable within that group but is capable of viciousness and deceit outside it. Within the group, I might be considered wonderfully trustworthy, but outside the group I might be thought utterly reprehensible in my abuse of any opportunity to exploit or harm certain others.

Ethnic bigots and my acquaintance of the past, if viewed strictly in the contexts of their own groups, might seem generally to be acting from trustworthy character or moral commitments. Seen in a different or much broader context they might seem to be not trustworthy either in character or moral commitments. On the account of trust as encapsulated interest, however, their actions might readily fit their own interests both within and outside their groups. We should therefore be clear which of these conceptions we are using when we attempt an explanation of some behavior.

If trust is grounded in encapsulated interest, then clearly it is, as noted earlier, relational. It is not merely a reflection of my character or yours. Little of the systematic empirical work on trust allows us to assess any relational elements. Much of the psychological work is on high and low trusters. Most of the survey work is on relatively loose claims of how much subjects "trust" most people or government. The game theoretic work often deliberately excludes any possibility that the players will have any broader relationship—often, for example, the adversary-partner of a player is unknown and will not be met again after a single, initial interaction. Claims from these bodies of empirical research therefore can tell us virtually nothing about trust as encapsulated interest or any other relational conception of trust. (I address many of the empirical studies later in relevant contexts.) Typically, the most we can get from this research is some insight into the readiness of people initially to take risks of cooperation with unknown others—usually very small risks. Because trust in our lives is generally relational and is commonly to be explained by relational considerations, one may hope that empirical studies will begin to take relational elements into account.

The discussion in this chapter is of trust between individuals. Later, in chapters 7 and 8, I take up generalizations from trusting

individuals to trusting groups or institutions. The goal in these chapters is to make sense of trusting groups or institutions in terms analogous to those of trusting individuals. When people say, in ordinary language, that they trust the government, *they do not mean anything closely analogous to what they typically mean when they say they trust another person*. That can become clear, however, only if we first unpack what ordinary individual-level trust is about in common instances.