

Trust: beliefs and morality

Jonathan Baron*

Trust is necessary for organizations to function well, both internally and in their relations with each other. Trust has been considered a form of human capital, a public good. Few if any social scientists or political leaders complain about there being too much trust. Many contributors to this volume are concerned with trust as a good.

Trust is important in the functioning of firms and institutions. One of the qualities of good academic departments, for example, is that people do their work for real, without pretending or excessive complaining.

Trust is also important in relations between citizens and government. Giving power to the government requires believing, or acting as if one believed, that the government officials involved will use the power correctly. Ultimately, this involves trust in one's fellow citizens. They must not only be trusted to do their jobs when they are government employees, but they must also be trusted to recognize and correct future errors of policy. Without such trust, change becomes difficult. People are afraid to risk giving power to the government for fear of some error that will not be corrected, so they attached to the status-quo.

Breyer (1993) has described a vicious circle resulting from mistrust of government: simplistic laws that tie the hands of government, government abuses as a result of trying to

*Citation: Baron, J. (1998). Trust: beliefs and morality. In A. Ben-Ner & L. Putterman (Eds.), *Economics, values, and organization*, pp. 408–418. Cambridge University Press.

apply those laws, more mistrust, more laws, etc. So trust in government may also require trust in the ability of government officials to make wise judgments. Of course, the officials must deserve such trust, and Breyer discusses ways to bring about a situation in which they do deserve it.

Trust is also important at an international level. We must trust people from other countries to negotiate honestly, to try to keep their commitments to treaties and other agreements, and not to use military force except in self-defense. International commerce clearly depends on trust in foreign business associates to be as honorable toward foreigners as toward compatriots. International trust will become more important as the world moves to solve environmental, population, and resource problems that affect many nations.

If trust is a good, and if we want to promote it, what exactly do we want to promote? In one sense, trust is a belief about the goodness or rationality of others. If we want person X to believe that others are good (or rational), and if X's beliefs about others are generally accurate, then either we must make sure that the other people *are* good or we must deceive X into having inaccurate beliefs. Of course, if X is cynical to the point of inaccuracy, then we can increase her trust by bringing her beliefs into line with the facts.

However, trust is not just a belief. It is also a behavioral disposition that is somewhat independent of belief. In this sense, we can make X more trusting by changing her behavior alone. We can promulgate a norm of trust as a behavioral disposition. Of course, such dispositions are affected by beliefs, and beliefs are affected by reality, but there is slippage at each of these links. Although the effort to influence beliefs independently of reality may require deception, the effort to influence behavior independently of beliefs requires nothing necessarily immoral or irrational.

We may thus speak of two kinds of trust. The first is believing something about others. The second is behaving as if you believed this. Although the second sense may be

derivative, it is no less real or important.

In this note, I discuss the nature of trust. Then I review some laboratory measures that could be interpreted as telling us about trust. Use of these measures may help to inform us about the psychology of trust, so that we can promote it better. I conclude with some comments about what sort of trust we ought to promote.

What we trust people to do

If we want to promote trust in society or in organizations, either as a belief or as a disposition, then we need to know more precisely what kind of trust we want to promote. Different kinds of trust correspond to different beliefs about the behavior of others.

There are several things we can believe about others. First, we can believe that others will not “tremble,” that is, that they will pursue their goals without error. This is a kind of belief in the competence of others. Second, we can believe that others are not myopic, that they are sufficiently concerned about their long-run self-interest. Third, we can believe that others will be sufficiently altruistic, that is, willing to sacrifice their own self-interest for some greater good. This category includes a belief that people are not sadistic or competitive in a way that leads them to rejoice in the losses of others.

Each of these dispositions might be limited in some way. In particular, it might be limited to some group of others, such as members of one’s firm, institution, or nation. Other classifications of the limits of trust are possible.

Of particular interest is the classification of trust in the altruism of others. We can have different beliefs about the conditions under which this altruism will be displayed. The altruism of others might be limited to members of a certain group, or to those who also have an opportunity to display the same kind of altruism in a way that might be seen as reciprocal (even when each person must decide without knowing the other’s decision). I shall call the last kind of situation “cooperative.” Cooperation is a response open to each

of several people, such that each cooperative response is better for this group on the whole but worse for the person who makes it. Trust in people's willingness to cooperate is especially important, for reasons I shall discuss.

Altruistic behavior – and hence trust in others' altruism – can also be classified in terms of the object of the self-sacrifice. Self-sacrifice can serve the interest of others, when some behavior involves a small sacrifice for self and a great benefit to others. Or, it can serve some principle, such as obligation to keep a promise, even when nobody benefits immediately. The principle in question might also be some form of fairness. Fairness might require one subject in an ultimatum game to give up half of her money, even though the other subject is much richer and thus benefits less from the money. Another principle that people might be trusted to follow is that of contingent cooperation, which specifies that you should cooperate when you believe that most others in your position will also cooperate (Baron, 1994). This might be called pitching in and doing your share.

Sometimes the various forms of trust are difficult to distinguish because the behaviors in question are motivated in more than one way. In organizations, for example, altruism and concern for long-run self-interest often work together to produce the same other-regarding behavior. So trust in long-term prudence and trust in cooperative altruism are indistinguishable. Workers might expect to be rewarded with promotions and raises for contributions they make to the organization. Often, though, the most effective workers seem to be motivated by a sense of duty or concern for others rather than their long-run self-interest, although they may occasionally think to themselves that their efforts will in the end be rewarded. Sometimes people are motivated by simple altruistic concern for their fellow workers. When people can be trusted to behave altruistically or cooperatively, many types of enforcement become superfluous.

Trust as morality

Trust is a virtue, not just a belief. The virtue of trust is the tendency to behave *as if* one believes that others will behave in the right way, and the tendency to value such behavior in oneself and others. A virtue is a kind of social norm that we endorse for others, and, by implication, for ourselves (Baron, 1996b).

Such a norm *could* take the form of endorsing self-deception. It could tell us to change our belief so as to think well of others. Although some people could interpret it this way, I do not. I think that the norm tells us to give others the benefit of the doubt *in our behavior*, whatever we might think privately. It tells us to have courage, to lie down and bare our necks, even when we are trembling inside.

A competing norm might be called the norm of suspiciousness. It tells us to watch out, to avoid entanglements, to have everything in writing, to hire a lawyer. People who violate *this* norm are seen as saps or suckers, as weak. If the norm of suspiciousness is a virtue, then it is a personal virtue, not a moral one. It does us little good to have a society filled with people like this. It benefits only the individuals, and, even then, only if others tolerate those who follow this norm. It is the kind of norm that parents might teach their children out of a concern for their children's welfare, but we don't hear preachers or university presidents telling us to watch our wallets.

The distinction between trust as belief and trust as a norm of behavior may seem somewhat artificial. People, after all, adjust their behavior to follow their beliefs, and they adjust their beliefs to be congruent with their behavior (e.g., Ellsworth & Ross, 1983). But not entirely. Moreover, thinking of trust as a norm leads us to think of different ways to increase it aside from deceiving people about what other people are likely to do, or aside from fomenting revolutions.

Following the norm of trust has an effect on both the beliefs and norms of others. It creates a virtuous circle opposite to the vicious one I described. My evidence that this

norm exists in society comes from my own experience. It was explicitly taught to me, over and over, by Leslie Cheek, the head of a school I attended in the eighth grade. He said that, if we act as if we expect the best from others, they will often behave better as a result. I have tried to follow this principle, when I remember it.

Use of games to measure trust

Various experimental games are sensitive to various forms of trust. The general form of these games involves a subject choosing between two options, call them trusting and cautious. The trusting option leads to a better outcome than the cautious option if the other subjects behave in a “good” way, and to a worse outcome if they do not. The choice of the trusting option thus suggests either a belief in the goodness of others or a willingness to behave as if one had this belief. The goodness may take the various forms I have sketched, from rational pursuit of self-interest to altruism. Only a few of the various kinds of trust have been explored in laboratory games, mostly those dealing with trust in the cooperativeness of others.

Berg, Dickhaut, and McCabe (1995) studied a simple trust game that took the form of an investment. In each round, one subject was given \$10. This subject could send some or all of that to a second subject, and the money would triple in value. The second subject could give back some money to the first. On average, the first subject invested about half of the money, and the second subject gave back about the same. The kind of trust at issue is cooperation, but a specific sort of cooperation in which the second player knew that the first had first behaved cooperatively to some extent.

Orbell and Dawes (1993; see also Orbell, Dawes, & Schwartz-Shea, 1994) studied trust with a prisoner’s-dilemma game in which each subject could choose to play or not. The payoffs for mutual cooperation or free riding were positive, but those for mutual defection or being a “sucker” were negative. The decision to play was thus a sign of trust

in the other player to cooperate. Cooperation was greater when the option not to play was available, mostly because defectors, expecting others to be like themselves, tended not to play.

In these kinds of games, it is important to ask subjects about their beliefs, which cannot easily be inferred from behavior alone. Subjects may, for example, cooperate because of a moral belief that this is the right thing to do, even if their belief that cooperation is in their self-interest would be insufficient to motivate cooperation on its own.

Evidence for attitudinal effects

Some evidence suggests that trust – in the form of trusting behavior – can be affected by experience with the behavior of other players in experimental games, as argued by Schotter in the present volume. This suggests that trust as a belief about others plays some role in trusting behavior.

Fleishman (1988) gave subjects false feedback about the behavior of others in order to manipulate their expectations. The effects of this feedback depended on whether the situation was described to the subjects as a “give some” game or a “take some” game. In the give-some game, each subject in a five-person group was given 100 points (each worth 1/3 cent) on each trial. The subject could give any number of those points to a central pool or stock. At the end of each trial, the number of points in the pool was doubled and divided among all the subjects. (Each subject would thus get 200 points per trial, if all contributed all their points.) In the take-some game, 500 units were given to the pool at the beginning of the trial, and each subject could take up to 100 units. Formally, the games are identical, except for the description in terms of giving or taking. Not giving a unit is the equivalent of taking it, and vice versa. To manipulate expectations about others, the experimenter told subjects after the first trial that the pool contained either 380 or 120 units. In the give-some game, subjects *gave less* on subsequent trials when they

thought that the pool contained 380 after the first trial than when they thought it had 120. That is, behavior was opposite from what others were perceived as doing. In the take-some game, they *took less* when they thought that the pool had 380. Behavior was the same as what others were perceived as doing. (These effects were small and should be replicated, but they are consistent with previous conflicting results cited by Fleishman.)

These results can be understood in terms of conflicting rules for behavior, which have been called fear and greed. Fear is fear of being a sucker. It induces behavior that is similar to the behavior of others, cooperating when others are perceived as cooperating and defecting when others are perceived as defecting. Greed is the tendency to take advantage of others' good will. Evidently, give-some subjects reasoned, "If nobody else is giving anything, then I'd better do it, but if others are giving a lot, I can take advantage of them." This is greed. Take-some subjects seemed to reason, "If nobody else is taking, then I'd better not take either, but if others are taking, then I can take too." This is fear. It seems to prevent harmful actions but not harmful omissions.

Other research has examined the continuity between laboratory measures and measures of attitudes about real-world analogs. Yamagishi and Sato (1986) found evidence for individual differences in fear and greed. On every trial, each of five subjects was given a number of points (worth money), some of which the subject could contribute to a common pool. All subjects received bonus points that depended on these contributions.

Two conditions were of special interest. In the *minimum* condition, the number of bonus points was determined by the size of the contribution of the subject who had contributed the *least*. Here, fear would be the major cause of defections: If you think that someone else is going to be a fink and give very little, there is no advantage to anyone from your not "finking out" yourself by giving nothing or very little. Greed would play only a small role here, because too much of it would penalize you as well as everyone else. In the

maximum condition, the bonus depended on the amount given by the highest contributor. This is like volunteering for a dangerous mission on behalf of a group, when only a few volunteers are required. Here, one is little affected by a few defectors, so fear of someone else's defecting plays little role. One need not fear that one's own cooperative action will be useless to the group, either. Greed, however, may play a larger role, for one's own contribution also matters little, except in the unlikely event that one is the highest contributor.

Yamagishi and Sato found that friends cooperated more than strangers in the minimum condition but not in the maximum condition. Presumably friends were less afraid that they would cheat on each other by giving less money, but friendship had no effect on personal greed. Subjects (Japanese students) were also given a questionnaire about their attitudes toward public questions, with items designed to assess both fear and greed as general motives in the subjects. Here are some "fear" items (pp. 69–70):

Help to developing nations should be limited to the minimum, because it is only exploited by a small group of people.

During the oil shock, people rushed to stores to buy a stock of toilet paper because people are concerned only with their own interest and not with the benefit of society.

There will be more people who will not work, if the social security system is developed further.

Here are some greed items (p. 70):

In order to be a successful person in this society, it is important to make use of every opportunity.

It is not morally bad to think first of one's own benefit and not of other people's.

The score on the fear scale correlated with defection in the minimum condition, but not in the maximum condition. In the minimum condition, where only one "fink" could prevent everyone from getting a bonus, subjects who expressed fear and distrust in the questionnaire were more likely than other subjects to defect. The score on the greed scale correlated with defection in the maximum condition, but not in the minimum condition. In the maximum condition, subjects who were greedy according to the questionnaire were the ones more likely to defect. These results not only indicate the existence of individual differences in the motives in question, but they also suggest that the motives are quite general, affecting attitudes toward public issues as well as behavior in laboratory experiments.

In the minimum condition, subjects' *expectations* about what others would do correlated highly with their own behavior; this was less true in the maximum condition. The minimum condition thus seems more sensitive to trust as a belief that others are cooperative.

In another study, Yamagishi (1986) used a short questionnaire to measure individual differences in trust. Subjects indicated their agreement with the following items, which are concerned with a variety of altruistic (vs. selfish) behaviors:

Most people tell a lie when they can benefit from doing so.

Those devoted to unselfish causes are often exploited by others.

Some people do not cooperate because they pursue only their own short-term self-interest. Thus, things that can be done well if people cooperate often fail because of these people.

Most people are basically honest (reversed item).

There will be more people who will not work if the social security system is developed further.

Each subject could contribute points to a public pool, where the points doubled in value and were then distributed among the four subjects. In some conditions, subjects could also contribute points toward a punishment fund to penalize the lowest contributor to the original pool, in proportion to the amount contributed to this fund. Subjects who were trusting, according to the trust scale, made higher contributions to the public pool when the punishment fund was not available. When the punishment fund was available, low-trust subjects contributed more to it than high-trust subjects, and the differences in contributions to the public pool disappeared. The low-trust subjects' contributions to the public pool were more affected by the availability of the punishment fund.

Parks and Hulbert (1995) used Yamagishi's trust scale to study individual differences in cooperation in public-good and resource-dilemma games. In the public-good game, three of five subjects had to "contribute" in order for all the subjects to get a bonus. On each trial of the resource dilemma, subjects could withdraw from a common pool, which increased by 10% after the withdrawals. Each game had a "no fear" condition, in which a payoff of zero was impossible. (In the public-good game, for example, contributors got their contribution returned if the bonus was not provided.) The trust scale correlated with contributions only when fear was present, in both games. The trust scale thus behaved like the fear scale of Yamagishi and Sato (1986), with which it shared a couple of items.

Note that the trust scale is ostensibly about beliefs. This does not mean that beliefs, rather than moral values, account for the correlations with behavior. People bring their beliefs into line with their values, as I noted earlier. So the correlation with behavior could be mediated partly by values.

These results say little about where the individual differences arise. Solodkin, Chaitas, and Baron (1995) examined individual differences as a function of national culture. We gave hypothetical social dilemmas to college students in Argentina, Mexico, and the U.S. For example, one dilemma read, “Imagine yourself living in a very polluted city, a city with so much pollution that many new laws have been made to fight it. One of these laws requires everyone to have their car inspected every 6 months, and if the emission levels are high, the car has to be fixed. Suppose your car fails the inspection, and imagine you have two alternatives: One is to pay \$100 to have your car fixed. This will not make it less likely that you will need it fixed after the next inspection, six months later. The other is to pay \$45 to the mechanic so that he gives you a permit to drive for six months.” After each dilemma, we asked subjects (along with other questions) what they would do, what proportion of people in the same situation would pay the \$100, whether they would be more likely to pay the \$100 if they thought that almost everyone else would do it, and whether paying the \$100 would be in their long-run self-interest.

Latins (Argentinians and Mexicans) did not differ from Americans (U.S. subjects) (or from each other) in stated willingness to pay the \$100. But Latins were less inclined to believe that *others* would pay it. This was so despite the fact that beliefs about others generally correlated with the subject’s own willingness to cooperate. There is evidently some other determinant of willingness to cooperate aside from beliefs about what others will do.

A hint about what it might be was in the answers to the question about whether cooperation was in one’s long-term self-interest. Subjects generally answered this affirmatively, but Latins were especially likely to do so. And the affirmative answer also correlated with stated willingness to cooperate. Thus, the willingness of Latins to cooperate may be decreased by their belief about others but increased by their belief in

long-run benefits. Unfortunately, we did not ask about the nature of these long-run benefits, but another study suggests that subjects confuse the benefit of their own cooperation with the benefit of everyone cooperating (Baron, 1996c). For example, in response to a scenario about overfishing, one subject said, “. . . by not cutting back on my current level of fishing I would more than likely be hurting myself in the future, for sooner or later the fish population would dwindle down to zero, leaving me with absolutely no income. I am willing to give up short-term financial growth for long-term financial stability.” Such a confusion supports a moral norm of cooperation.

Other effects of trust

If trust in others is a norm, what form should this norm take? Should we act as if we believed that people are perfect utilitarians? rational pursuers of their self-interest? contingent cooperators?

I suggest the following norm as something worth encouraging. We should think of each other as willing to sacrifice somewhat, but not excessively, for the common good. Thus, when setting up cooperative agreements, some enforcement mechanism may be needed, but it need not be Draconian. Spot audits of taxpayers together with small penalties for violations may be sufficient, for example. We should also think of each other as willing to cooperate to set up such reasonable enforcement mechanisms, that is, to solve second-order social dilemmas. We should endorse such willingness as a social norm, and we should behave as if we expect it from each other.

Norms of good citizenship should also include trust in government and in the ability of democratic government to correct errors. We should discourage arguments based on mistrust in government, especially when these arguments are used to oppose an otherwise good solution to some social problem.

Norms of this sort could be taught as part of civics in school. They could be

supported by politicians, journalists, and professors. Such support could work against the kind of vicious circle that leads to waste and disorder.

References

- Baron, J. (1994). *Thinking and deciding* (2nd ed.). New York: Cambridge University Press.
- Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, 8, 71–83.
- Baron, J. (1996a). Political action vs. voluntarism in social dilemmas and aid for the needy. (Unpublished manuscript, Department of Psychology, University of Pennsylvania).
- Baron, J. (1996b). Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy*, 12, 165–182.
- Baron, J. (1996c). The long-run fallacy as a reason to cooperate in social dilemmas. (Unpublished manuscript, Department of Psychology, University of Pennsylvania).
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social-history. *Games and Economic Behavior*, 10, 122–142.
- Breyer, S. (1993). *Breaking the vicious circle: Toward effective risk regulation*. Cambridge, MA: Harvard University Press.
- Ellsworth, P. C., & Ross, L. (1983). Public opinion and capital punishment: A close examination of the views of abolitionists and retentionists. *Crime and Delinquency*, 29, 116–169.
- Fleishman, J. A. (1988). The effects of decision framing and others' behavior on cooperation in a social dilemma. *Journal of Conflict Resolution*, 32, 162–180.
- Orbell, J. M., & Dawes, R. (1993). Social welfare, cooperators advantage, and the option of not playing the game. *American Sociological Review*, 58, 787–800.
- Orbell, J., Dawes, R., & Schwartz-Shea, P. (1994). Trust, social categories, and individuals: the case of gender. *Motivation and Emotion*, 18, 109–128.
- Parks, C. D., & Hulbert, L. G. (1995). High and low trusters' responses to fear in a payoff matrix. *Journal of Conflict Resolution*, 39, 718–730.

Solodkin, A., Chaitas, S., and Baron, J. (1995). Students' attitudes toward social dilemmas in Argentina, Mexico, and the United States. (Unpublished manuscript, Department of Psychology, University of Pennsylvania).

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.

Yamagishi, T., & Sato, K. (1986). Motivational basis of the public goods problem. *Journal of Personality and Social Psychology*, 50, 67–73.