Trust as an Action

Paul Dumouchel
Ristumeikan University
Core Ethics and Frontier Sciences
56-1 Toji-in Kitamachi, Kita-ku, Kyoto 603 8577
JAPAN
Dumouchp@gr.ritsumei.ac.jp
Many scholars view trust as an expectation concerning the future action of other agents. For example, in *Trust a Sociological Theory* (1999:25), Sztompka defines trust as “a bet about the future contingent actions of others.” Sztompka insists on the fact that this bet, or expectation, is trust only if it has some consequence upon the action of the person who makes the bet. Suppose that we are watching a baseball match and I tell you: “I bet the pitcher is going to throw one of his famous curve balls.” I am clearly making a bet about the pitcher’s future action, but this expectation is not trust because it has no consequences whatsoever for my own future action. “Trust”, Niklas Luhmann wrote, “is only involved when the trusting expectation makes a difference to a decision.” (1979: 24) Similarly, when Diego Gambetta attempts to summarise the different conceptions of trust in the volume that he edited on that topic, he tells us that “trust is a particular level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action… and in a context in which it affects his own action.” (1988:217) Gambetta, like Sztompka and many others, views trust as an expectation concerning the future action of some other agent(s) in a context where that expectation has an influence upon the action of the person who has the expectation. The list of authors who view trust in this way could be extended, but these three should be sufficient for our purpose. I propose to call “cognitive theories of trust” theories that construe trust as an expectation about the future behaviour of other agents. These theories are cognitive in two ways. First, they consider the uncertain, incomplete and often probabilistic knowledge that an agent has of the future action of another to be the central element of trust. Cognitive theories of trust also recognise a non-cognitive element in trust: that the knowledge, expectation or subjective evaluation must be related in some way with the agent’s decision to act. Sztompka argues that this second element introduces an active, social and objective dimension to trust. As we have seen, we should speak of trust only if the knowledge element it contains leads to action or at least to a decision. Nonetheless, this second element also reasserts the importance of the cognitive dimension. It says that what explains the agent’s action is the expectation. The uncertain and imperfect knowledge that an agent has about the future behaviour of others occupies a central place in the complex of beliefs and desires that explains the agent’s action and that either causes or motivates him or her to act. Such theories are also cognitive for another reason. They consider that what makes trust necessary is that we have imperfect knowledge of the world and especially of the future behaviour of other persons. According to them, if we knew in advance how others would act we would not need to trust. They see trust as a means compensating for this lack of knowledge. From this point of view, the role of trust is cognitive.
inasmuch as it replaces the knowledge that we do not have and allows us to act as if we had that knowledge.

Despite of their success in many disciplines, including sociology, economics and philosophy, cognitive theories misrepresent trust in many ways. Imagine that we are playing tennis together. I have just hit the ball, sending it to the far right-hand corner of the court. I now rush to the left end of the court, where I believe your backhand return will land. You are faster than I thought. Having seen or perhaps anticipated my movement, you place the ball at the right end on my side of the court. I have just lost the exchange. It seems strange to say that when I ran to the left end of the court I trusted you. Yet this is what the cognitive definition entails. I had an expectation about your future action and acted accordingly. No one would think that as a consequence of this exchange, I have now lost trust in you or that I believe you to be untrustworthy. Clearly, cognitive theories of trust cast their net too wide. There are many circumstances in which action is guided by expectations about the behaviour of other agents, most of which do not involve trust. Think of conventions, games of co-ordination in general and innumerable everyday situations. Consider, for example, the following fictive dialogue. You have just asked me why I did not phone to let you know I was coming back earlier:

- I did not phone because I expected you to be out when I left. At 4:00 PM on Wednesday you usually have your driving lesson.
- …
- Oh! You decided not to go today.

Neither trust nor distrust is involved in my decision not to phone or in my discovery that my expectation was wrong. Cognitive theories of trust do not contain any criteria that allow us to distinguish trust from other actions that we undertake on the basis of expectations about the behaviour of other agents. It is therefore not surprising that in cognitive theories of trust, the specificity of trust often tends to be lost. Trust comes to be equated with some ill-defined “social capital” that is essential for co-operation,1 or with moral sentiments in general,2 or with whatever it is that constitutes the social bond.3 This is true even of authors who go to great lengths to distinguish trust from other apparently closely related phenomena like confidence and familiarity.4 However, this lack of specificity and inability to differentiate trust from other forms of behaviour that rest on expectations about the behaviour of other agents is not the only or even the main problem of cognitive theories of trust.

---

2 For example: M. Hollis (1998).
3 A. Seligman (1997).
Conceived as a subjective probability or expectation, trust is more or less identified with a judgement about the trustworthiness of another agent. It is this judgement that explains the agent’s decision to act. The second difficulty that cognitive theories of trust face is empirical evidence suggesting that this is not the case. It is not the way agents act. In *Autonomy and Trust in Bioethics*, Onora O’Neil argues that agents very often trust people whom they think untrustworthy.\(^5\) It is strange behaviour to be sure, but, when we think about it, perhaps not infrequent. How should it be explained? One possibility is to suggest that such agents are irrational and motivated by some obscure force to act against their better judgement, which recommends that they not act co-operatively. However such agents are incomprehensible. This way of dealing with the anomalous evidence is adequate only if irrational agents, or circumstances in which agents act irrationally, are rare. Otherwise our explanatory theories lose their utility. Another possibility is to suggest that it is the evidence itself that is untrustworthy. It could be argued that even though people sometimes say that they place their trust in individuals whom they think are untrustworthy, this should not be taken at face value. Such people will usually add that they do this because they are forced to or because they have no choice. What they mean is not that they act against their better judgement but that if circumstances were different they would have acted in another way. Their level of trust may be low, but that is not the only thing that explains their decision. Theorists usually argue that when everything is taken into account, such as the cost involved in not acting co-operatively, the expected benefits, the window of opportunity, and so on, the agents’ decision to trust is not irrational.

Yet, as Gambetta suspected, this conventionalist strategy for avoiding anomalous evidence threatens to make trust irrelevant. At first sight, it seems highly reasonable to consider trust to be only one element in the decision-making process. In order to understand an agent’s decision to trust we should take these other elements into account. When the stakes are low and the probability is high that the other person will respond co-operatively, for example, that you will bring back my book, perhaps because I am your professor and it is in your interest to do so, I might trust you even if I think that you are not a very trustworthy person.\(^6\) Clearly trust is not the only thing that explains our decisions to act co-operatively; other elements enter into the process. However, making this insight operational in our model of rational choice requires that trust be convertible into the currency of any other element that determines the issue. This means that less trust can always be compensated by more interest, as in the above example, or conversely that more trust is necessary when interest is a weak incentive. When the desired outcome me is highly probable, trust has a smaller role to play.


\(^6\) This is essentially the structure of Russel Hardin’s “encapsulated interest” theory of trust, in Hardin (2001).
and when it has a low probability, more trust is required in order for a person to act co-operatively. All of this is very commonsensical, but the consequence is that in the decision-making process, trust can always be replaced or compensated for by more interest or by manipulating the environment in one way or another. In theory at least, trust is never necessary. The goal of Gambetta’s (1988) paper, “Can we trust trust?”, is to argue that, in real life, trust cannot be replaced as easily as it can be in our theoretical models.

Actually, the difficulty is more serious than Gambetta thought. It is not only that in such a model trust can always be replaced or compensated for by any other element that enters into the decision-making process, but that trust and distrust cannot explain anything: neither a decision to act co-operatively, nor a decision to defect. According to this model, to trust in the sense of placing trust, is to act co-operatively on the basis of an expectation concerning another agent’s future action. Yet, as Gambetta noted (1988; 222), the decision to act, i.e., to place one’s trust, must be independent of any specific level of subjective probability concerning the other agent’s action. That is to say, there cannot be a probability threshold below which the agent always decides not to act co-operatively, or any specific probability that is identified with trust. This is precisely because trust can be replaced by any other element in the decision-making process. In cases where our interests lead us to co-operate, trust is not as important and a lower level of subjective probability will be enough for me to co-operate. On the contrary, when the interests of agents diverge sharply and the stakes are high a greater level of subjective probability will be required in order for me to trust. However as Gambetta himself reminded us in the first paragraph of this paper, cognitive theories of trust identify trust with the subjective probability that we use to assess whether an agent will perform a specific action. It follows that to say that the decision to trust is independent of any level of subjective probability, as Gambetta does (1988; 222), is to say that the decision to trust is independent of trust! Cognitive theories of trust define trust as an expectation concerning another agent’s action that is relevant to the decision to act. However the decision model to which they resort simultaneously requires that the action of the agent who has the expectation be independent of that expectation!

Finally, viewed as an expectation, trust is very strange indeed because in many cases it is an expectation that cannot be represented, either by the agent who has the expectation or by the theorist that attributes it. That is to say, it is not possible to assign any specific content to the expectation. It is easy to come up with examples where the language of expectation apparently works. “You told me you would be at the airport at 1:00 PM. I trusted you and on the basis of that expectation I was there on time to meet you.” Here the language of expectation works because we can give a definite content to the expectation, namely, “that you will be at the airport at 1:00 PM”, and it is possible to assign to it a specific probability. But in many cases, and perhaps in all the cases where trust is fundamental to a relationship,
this is not possible. Trust is usually considered a central element of most types of long-term partnerships. However, long-term partnerships are by definition turned towards an unknown and unknowable future. The meaning of trust in such contexts is that when faced with unexpected circumstances our partner will act co-operatively, though we do not know how he or she will act. How can we represent the future contingent action of another if it takes place in an open-ended future that is unknown to us? How can we assign a subjective probability to such an action? What is the content of: “He or she would never do a thing like that”? What is “a thing like that”? How do you represent “that”? Such statements make sense only if there is a previous statement that has determined the action that falls into the category of “things like that”. Yet it is precisely against the ill-defined category of “things like that” events that trust offers protection.

Let us summarise our efforts so far. Cognitive theories of trust offer a definition of trust that lacks specificity and does not allow us to distinguish trust from other actions that we undertake on the basis of expectations concerning the future actions of other agents. Such theories also face anomalous evidence, namely, the fact that we often place trust in agents who we believe to be untrustworthy. In their effort to accommodate this evidence cognitive theories resort to a decision model in which trust cannot explain the agent’s decision to act, and distrust cannot explain his or her decision not to act. Consequently the agent’s decision must be seen as independent of trust, yet trust is defined as an expectation that is relevant to the agent’s decision to act. Finally, in many cases trust cannot be represented as a definite expectation with a specific content. The language of expectation then becomes metaphorical to the point of absurdity. What does it mean to say that my action is explained by an undefined expectation concerning an unknown future?

Faced with these difficulties two routes at least are open to us. One is to abandon cognitive theories of trust, as I will soon argue that we should, the other is, so to speak, to “bite the bullet” and to concede that trust cannot be distinguished from any other type of cooperative phenomena. This last option is pretty much what Michael Bacharach and Diego Gambetta propose in a recent article “Trust in Signs” (2001). Trust, they claim, is not in its essential sui generis; it is “a complex phenomena (sic) but one whose main elements are met in other domains of decision making.” (2001; 175) In other words, there is nothing specific about trust; the word does not single out a homogeneous class of decision problems or of cooperative endeavours. Nonetheless, they, somewhat surprisingly, argue that the problem of trust does not in consequence disappear; rather, it is displaced. The problem of knowing whether or not an agent will cooperate in a given situation is what our authors define as the primary problem of trust. They take it for granted in the model they propose that this problem is solved and advocate a move to what they call the secondary problem of trust. That secondary problem is that of assessing whether the signs agents give to indicate that they will
cooperate are reliable or not. In other words, the central problem of trust is not of knowing with what probability an agent will adopt a non-dominant strategy in a game of imperfect cooperation, but that of knowing what type of game we are playing, one where the other player’s dominant strategy is cooperation or one where it is defection. More precisely the problem of trust, according to them, is that of knowing if the other agent is trying to deceive us by signalling that he or she is playing a game of one type when in reality the agent is engaged in a different type of game.

The real difficulty we face in life, according to them, is the reliability of signs. This move to the secondary problem of trust, they argue, opens up a coherent domain of inquiry that is common to both human and animal society. “The problem of the truster with respect to trustworthiness is the problem of the employer with respect to productivity, of the dog with respect to the belligerence of the hissing cat, and of the predator with respect to the toxicity of the striped worm.” (2001: 174) In short, our authors propose to replace the problem of trust with the problem of deception understood as being induced by an unreliable sign to perform an action one would not otherwise have undertaken. However, as the examples mentioned above suggest, the problems of trust and of deception do not really correspond to each other. That is to say deception is a phenomenon with a much wider range than misplaced trust. As various authors have argued, deception can be to everyone’s advantage, among other reasons because a threat that is not acted upon because it is believed can leave every player involved better off independently of whether or not the threat would have been honoured, that is to say, even if some players have been deceived. Understood as the question of the reliability of signs, the secondary problem of trust reaches far beyond what is usually understood by the question of trust. Of course Bacharach and Gambetta can claim that this is not a difficulty. Given that trust is so ill-defined and hard to pin point it is not surprising that a scientific theory will encompass problems and questions that are not captured by the everyday term.

Perhaps is this true. However, as John Maynard Smith and David Harper (2003) recently argued, in biology the question of signs among animals is not about deception but about the fact that signals are on average reliable. The question is not to determine which

---

7 When they present their model (2001:150-152) the authors distinguish between the players “Raw-Payoffs” and their “All-in Payoffs”. The first ones they tell us are motivated by “simple self-interest” while the others are the players’ Payoffs “all things considered”. Given the values they assign to these, when Raw Payoffs are taken into account, defection is the dominant strategy for both trustee and truster but when they are replaced by All-in Payoffs cooperation is the dominant strategy for both players. By definition, these different payoffs define two different games. Therefore the problem becomes one of knowing which game we are playing: one that is defined by the Raw Payoffs where cooperation is impossible or one that is defined by All-in Payoffs where cooperation is the only rational solution. It immediately follows that the “All-in Payoffs” game postulates that the primary problem of trust simply does not arise. The secondary problem of trust is then that of knowing if the signs that agents give to indicate that they are trustworthy, i.e. that they are playing a game defined by their All-in Payoffs, and reliable, given that agents have an incentive to lie.

signals are reliable and which are not but to understand why so many of them are reliable.

“Why are animal signals reliable? This is the central problem for an evolutionary biologist interested in signals. Of course, not all signals are reliable: but most are, otherwise, receivers of signals would ignore them.” (2003: v) Clearly Bacharach and Gambetta follow a different agenda and in spite of their claim to the contrary there does not seem to be much difference between the primary and the secondary problem of trust. “A signal”, they say, “is an action by a player (the “signaller”) whose purpose is to raise the probability that another player (the “receiver”) assigns to a certain state of affairs or “event”.” (2001: 159) Given this, it seems hard to avoid the conclusion that trust in signs is an expectation concerning a future state of affair and that this theory is just another cognitive theory of trust. It is therefore difficult to see why this recent proposal should escape any of the difficulties listed above. In fact it does not. Signs of trustworthiness are only one of the elements that enter into the decision to act cooperatively. Their absence can be compensated by the fact that the interests of the players converge strongly.9 In other words, there is no other specificity to the problem of signs regarding trust, than that which is postulated by our authors as a premise of their model. There is no secondary problem of trust that is distinct from the primary problem.

In view of these difficulties and in spite of such theories’ success in many disciplines, I think that we should abandon cognitive theories of trust. Instead of viewing trust as an internal cognitive or psychological element that explains or motivates an agent’s action, I propose to start from the action of trusting itself, i.e., from the characteristics of trust as an action. Unlike expectations or psychological dispositions, actions can be observed in the world. I think that there are two characteristics of trust as an action that are particularly important here. The first is that trust comes into play only in situations where the interests of the agents partially diverge and partially converge. In the relatively rare circumstances where the interests of the agents either perfectly converge or diverge, trust has no place. In zero-sum games and in games of pure co-ordination, trust is neither necessary nor useful for reaching equilibrium points. All that is needed is knowledge of our own interest. In circumstances in which interests partially converge and partially diverge to trust is to opt for convergence, i.e., to chose co-operation. But, as we have already seen, this is insufficient to distinguish trust from other types of co-operative action in circumstances where there is imperfect convergence of interests. I prefer to cross the intersection before you because I am in a hurry, but I would rather let you pass before me than crash into your car. If that happened, I would get at my meeting even later. Choosing the co-operative strategy of yielding the right of way does not require trust, simply the knowledge of the speed at which you are coming.

The particularity of trust as a mode of co-operative action is that by trusting, a person makes himself or herself vulnerable to the agent who is trusted in a way that would not exist had the person refrained from trusting. To trust is to act in such a way as to give another agent power over us. In other words, when I trust I increase my vulnerability to another agent through an action of my own, and that action is precisely what trust is. If I had not acted I would not be vulnerable, or at least not as vulnerable to the other agent. For example, when I trust you with an important secret, I give you power over me, power that you did not have before. The same is true when I pay now for something that you will deliver tomorrow. I give you means of harming me that you would not have if I had decided not to pay before delivery. I trusted you. To trust is to act and not simply to expect because it is the act, not the expectation, that gives the other agent power over the person who trusts.

To trust is therefore to act in a very special way. It is to act in such a way that as a result of one’s action another agent gains power over us. In general we do not act in that way in order to give another agent power over us. I confide in you because I need support and help. I pay before delivery in order to save 10%. In neither case do I act with the intention of giving you power over me. That is to say, giving you power over me is not my goal. Rather, it is something that I also do while doing something else and that I need to do in order to do that something else. I cannot obtain your help and support unless I tell you my shameful secret. I cannot get 10% off the price unless I pay before delivery. This is not true of every type of co-operative behaviour. In a convention, for example, it is instead when I do not act in accordance with the convention, i.e., when I act non-co-operatively, that I expose myself to sanctions. By acting in agreement with a convention I do not give the other parties in the convention power over me.

It is important to see trust as action, e.g., an act of my own by which I give someone else power over me. It is not simply that as a consequence of my action you have gained power over me. Though this situation can also exist, as such, it does not involve trust because the consequences of my actions are various and many are unintended. Unaware of the sharp shooter’s presence, I stepped out into the open and gave him clear shot. It is not because I trusted him that I died, but because I was ignorant of his presence and perhaps imprudent. To say that trust is an action means that even if it is not my intention to give someone else power over me it is nonetheless intentionally that I do it. I do not take a taxi with the intention of giving the driver power over me, for example to drive me all over town before bringing me to my destination. Yet, I intentionally put myself in his power. I trust him. The distinction between doing something intentionally and doing something with an intention dates from Anscombe and there is nothing mysterious about doing two or more actions simultaneously. I can both walk in the rain and go visit you, and though it is not my intention to walk in the rain, but to visit you, it is certainly intentionally that I walk in the rain.
In the case of trust, unlike that of walking in the rain, the action requires that the other agent to whom it is directed recognise it as such. Because we are in the domain of human interactions, actions are what they are only to the extent that they are perceived as what they are. Some have argued that the most perfect gift is one given when the recipient not only does not know who the donor is but is also unaware that what he has found has really been given to him. Others may think that such a “gift” does not deserve to be called by that name. Whatever the case may be, all agree that giving when the action remains unknown to the recipient is a different action from a gift that is public and recognised as such because it has different consequences in the world. Trust is a specific action that has some very particular consequences. Among these is the creation of a bond. When I trust someone I impose a form of obligation upon him or her, namely the obligation not to betray my trust. This obligation exists only if the other recognises in some way that I have given him or her power over me.\(^{10}\) This means that trust transforms the dynamics of the situation. If I make myself more vulnerable to you, I obtain in exchange a claim upon your future action. If you break my trust, I can reproach you for your action. In a prisoners’ dilemma, I cannot reproach you your defecting unless I have first trusted you. Your action is just what should be expected. Trust creates a normative expectation in the sense that Niklas Luhmann gives to that expression.\(^{11}\) An expectation is normative if, when it is disappointed, the direction of correction is not to change the expectation, as in the case of a descriptive or cognitive expectation, but to attempt to change the world. Even if I was wrong to believe that you would honour my trust, I still believe that trust should be honoured. Cognitive theories of trust propose that we trust because we have descriptive expectations about the future behaviour of other agents. From the point of view of trust as an action, it is because we trust that we have a normative expectation.

Trust is a very important type of action that plays a fundamental role in our lives. There are many different circumstances where we must put ourselves in the hands of others. We are brought to give doctors and lawyers power over us for example, but also to place ourselves at the mercy of friends and lovers, colleagues at work, university administrators, research consultants, stockbrokers, civil servants, travel agents and, yes, of course, taxi drivers. We can at times refrain from trusting, which usually entails foregoing some opportunity, but we cannot completely avoid trusting. Cognitive theories of trust consider that what makes trust necessary is our limited knowledge of the world and especially of the future contingent actions of other agents. Viewing trust as an action suggests that what makes trust necessary is not so much our lack of knowledge about the world as the fact that we depend

\(^{10}\) Some authors like Bacharach & Ganbetta (2001:161) argue that we look people in the eyes to discover signs of whether we can trust them or not. I think we look at them in the eyes because we want them to acquiesce that they have been trusted and to recognize the responsibility that goes with it. Put another way, we are not trying to learn something about them, we are trying to tell them something.

\(^{11}\) N. Luhmann (1985:31-40).
upon each other. It is true that we do not have perfect knowledge of the world, but even if I did know in advance what you will do, it does not follow that I could always avoid giving you power over me. A child who is confronted with an abusive parent or educator is often forced to give power to a person whom he or she knows will take advantage of it. What pushes him or her to do this is not lack of knowledge about the future behaviour of the adult but the child’s relative helplessness. What motivates a person to give another power over him or her is that the trusting agent believes that some good will follow from it. That good may be large enough to compensate for the evil the agent also knows will follow from the action. When cognitive theories try to accommodate this fact, they are lead to the absurdity that trusting (the action) is independent of trust (the expectation or moral sentiment).

When trust is construed as an action rather than as a sentiment or internal state, such absurdities are avoided. The specificity of trust among different forms of co-operative behaviour is maintained. The evidence that we often place our trust in agents whom we “do not trust” no longer appears anomalous. Finally, it suggests an interesting avenue of inquiry into the relationship between trust and various forms of political relations.

Bibliography


---

1 Previous versions of this paper have been read at Kinjo University in Nagoya, at the University of Kyoto and at the 48th Congress of the Canadian Philosophical Association held at the University of Winnipeg. I wish to thank the participants at those meetings and others for useful comments and criticism, in particular, Martin Aiken, Mary Baker, Rolf George, Harcel Hénaff, Steven Lukes, Mathieu Marion, Kenneth Nickel, Masachi Ohsawa, Matthew Taylor, Hiromi Tomita and Kozo Watanabe.