

Draft chapter: Is Trust Obsolete?  
(to eventually appear in *The Cost of Honesty*)

Judith Donath, April 19, 2021

Trust has been essential for the development of human societies. It provides the confidence that enables us to interact with others without needing a continual stream of assurances that we will not be cheated; it allows us believe what others tell us without laborious confirmation. To be amongst people we trust is relaxing: we eat food without fearing it was poisoned, confide secrets knowing they won't be betrayed, leave valuables about certain they won't be stolen. Yet with the rise of surveillance and other confidence-ensuring technologies, the role of trust is changing.

By many accounts, trust is declining ((Putnam 1995; Xin and Xin 2017) ). This trend is often seen as a sign of societal decay, the rational if regrettable response to an increase in other people's untrustworthiness, dishonesty and unreliability (Rainie, Keeter et al. 2019)). Articles with headlines such as "Collapsing Levels of Trust are Devastating America" see cause for distrust everywhere, from the media's presentation of a distortedly bleak and violent view of the world to single-parent households and loosening gender identities. And, while some of that blame is scatter-shot and illogical<sup>1</sup>--and there is no definitive evidence that people have actually become less trustworthy--there certainly are significant cultural changes, such as growing wealth disparities, that contribute to trust's diminution.

An alternative or additional hypothesis for trust's decline is that the need for it is diminishing. Growing numbers of institutions and technologies are taking over the task of ensuring that interactions go smoothly, people cooperate, and obligations are upheld. Think, for example, of buying on credit: to do so once required earning the store proprietor's trust through a perhaps lengthy relationship-building process; today, we

---

1. This and similar articles have much rhetorical energy, but are weaker on rigor and logic, often conflating cause and effect, and suffer from the same logical inconsistencies. For instance, Brooks cites fluid identity as a cause of declining trust, seems to tout 1950s homogeneity as a somewhat idealized time of high trust — yet also cites the reason younger people are less trusting is their concern that Americans do not respect people who are not like them — a concern that a return to a more stifling social uniformity would hardly assuage.

simply wave a card,<sup>2</sup> and, in a process devoid of human interaction, distant banks authorize the transaction. Or imagine a manager overseeing a fleet of drivers: whereas they once needed to trust a driver's account that traffic, not a personal detour, caused a late delivery, today location tracking-devices enable them to oversee the fleet with impersonal accuracy. With the increasing ubiquity of such confidence-ensuring innovations, the need to rely on trust withers.

Decreasing trust would seem to be an unambiguous loss. The connotations we have with trust are positive: to be trustworthy is a virtue, and to live and work with people you trust desirable. A "trust-free society" sounds Orwellian, or simply oxymoronic. Yet both the value of trust and the outlook for a future in which it plays a diminished role are more complicated than they may seem.

Yet, it is possible to argue that diminished reliance on trust is beneficial. Trust, far from being an unfettered ideal, is problematic. We can be poor judges of trustworthiness, and then suffer the sometimes severe consequences of misplaced trust. Trust is often based on personal relationships and in-group connections; it thus encourages homogeneity and exacerbates bias. And, trust does not scale indefinitely; as our societies grow from local villages to global networks, trust can no longer be relied upon to reliably facilitate cooperation. In other words, trust, though we hold it high esteem, can be unreliable, is antithetical to diversity, and seems increasingly anachronistic in a globally networked world. In thinking about what sort of world we would like to live in it, we should not blindly trust the goodness of trust.

First, some definitions.

Constraints, sanctions and trust are all ways of achieving confidence. If I am leaving the house, I want to be confident no one will not break in during my absence. How? I could use constraints, such as deadbolts and heavy window gates, to make it impossible for them to do so. I could deter burglars with a credible threat of arrest, by installing a police-alerting alarm system. Or, I may trust the members of my community not to break in, either because I believe they are not the sort of people who would do so, or because we had established a relationship I believe they valued enough to not jeopardize.

Confidence measures your assessment of how likely it is that something is true, of the veracity of some (possibly implicit) claim. The choice to accept the claim and change your

---

2. The physical credit card is becoming obsolete and will eventually be replaced by some form of biometric identification.

beliefs or behavior (ranging from a subtle shift in your impression of another to a major life decision) is a function of your confidence in the claim and your perception of the risk that being wrong entail<sup>3</sup>. Ultimately, it is confidence that makes our interactions possible; trust is one way of achieving confidence, but it is not the only way.

Constraints provide confidence by making the undesirable behavior impossible. If I do not trust the people in my neighborhood, I can use deadbolts and heavy window gaits to constrain entrance to my house; with them, I am confident my house will not be burglarized in my absence. Intrinsic signals [the subject of an earlier section of this book] constrain communication to honest claims about one's abilities, resources or intentions by being impossible to produce deceptively. It is an intrinsic signal of strength to lift a heavily weighted barbell: no matter how much I want to convince you that I am strong, only if I truly am will I be able to get it off the ground. An important feature of constraints is that they are often self-contained: they do not require a larger social structure or institutions.

Another way of achieving confidence is through sanctions--punishments that make the undesirable behavior unprofitable. To be deterred by sanctions, one needs to know both that the sanctions exist and that one is likely to be caught. While the latter is not a problem in the case of immediate retaliation, much sanctioning among humans is carried out by third parties, making surveillance an integral part of sanctioning. Like the locks, an alarm system connected to the police station can give me confidence that my house will not be burglarized. Sanctioning is an effective deterrent if one believes that the probability of being caught (the power of the surveillance apparatus) is high enough in combination with the costliness of the sanctions and the likelihood that they will be enforced so as to render the misbehavior unprofitable. The word "believe" is important here. If you believe that an omniscient spirit is watching you and will torment you for eternity if you violate its rules, you will be motivated to act in accordance with those rules, regardless of the objective reality of that spirit or its punishments. More prosaically, the converse: if you are unaware of a surveillance and sanctioning system, you will not be deterred (though you may well subsequently be caught). The surveillance aspect of sanctioning is of particular interest here, for new technologies—the internet, ubiquitous cameras, face recognition, etc.—greatly increase the observational reach and capabilities of sanctions, and thus potentially diminishing the need for trust.

Trust supports confidence by making honesty and cooperation more rewarding — it

---

3. My confidence in my ability to jump 10 feet may be plenty if I am deciding whether to jump across a shallow puddle, where the risk of failing is wet feet; it may not be enough to get me to jump across a deep chasm of the exact same breadth.

motivates us to act in the interest of another. Like many of other behaviors we discuss, there is a neuro-chemical underpinning to trust—we, or at least most of us, have some natural inclination to trust and be trusted.

Trust is notoriously hard to define, so much so that entire papers have been written about the confusion of its conflicting definitions (McKnight and Chervany 1996). Here, I propose a definition of trust based on its evolutionary roots as a motivator of prosocial behavior, one that frames trust as a relationship quality, a feeling one has about another. This definition highlights the subjective experience of trusting and being trusted, which many functional descriptions omit. When it is important to distinguish this meaning of trust I will use the term *affective trust*.

(Affective) trust is the positive feeling you have about another based on the belief that their intentions towards you are good, due to the value they place on their relationship with you and/or their general good morals and intentions.

This is similar to the definition of trust proposed by political scientist Russell Hardin, who has written extensively about trust. Hardin defines trust as “encapsulated interest”, wherein “I trust you because your interests encapsulate mine to some extent—in particular, because you want our relationship to continue.” But Hardin does not include any emotional element in his definition of trust.

The emotional component of trust is important. It is an inherent part of the experience of trusting another, of being trusted by another, and of acting in a trustworthy way. Its existence explains otherwise puzzling observations about our decisions about when to cooperate and our reactions to having our trust betrayed. And on the other side, being distrusted, distrusting another, acting untrustworthy—even knowing that someone thinks you are untrustworthy—all elicit vivid emotional response.

That trust involves emotion is recognizable from our everyday experience. Being entrusted with increasingly valuable things feels different than having your credit card company inform you that they have raised your credit limit. The latter might be quite handy, and you may be happy to have that added flexibility, but it is not the feeling of being trusted. The feeling that accompanies trust is not imaginary; it is an affective response whose physical manifestation can be seen in one’s neuro-chemical responses. While our understanding of these affective/cognitive systems is still incomplete, understanding the work that has been done helps us understand the range of behaviors connected to trust, including the trust we extend to those we see as part of our own group, and the distrust we have for outsiders. As we evaluate the consequences of replacing trust with other technologies and institutions as a means of gaining confidence, we need to recognize and value trust as an experience itself, and not only a means to an end; this is crucial if we are to

truly weigh what we gain and lose in this shift.

Trust takes work. He *earned* her trust. Sometimes, the process is explicit. A parent might start off with a new babysitter by having her come over to watch the baby while the parents are home but busy. Then, they will have her watch the baby while they are out for an hour, then for longer periods. Eventually, they trust her enough to leave the toddler with her while they are away on a weekend trip.

Discussions of trust often mention risk. Sociologist Diego Gambetta wrote ““For trust to be relevant there must be the possibility of exit, betrayal, defection”. This is especially true for building trust: if I don’t see you in a position where you could reveal yourself to be untrustworthy but do not, how would I know that you would not do so? Gifts and favors given and received at staggered intervals build trust in a relationship through reciprocity, and the underlying risk of forgetting or defecting from the chain.

The sensation of risk taking can trigger the sense of bonded togetherness that trust fosters. Imagine spending an evening talking with a friend. First, you chat about the weather, or the mundane activities of mutual acquaintances. But then the conversation slips into illicit gossip territory—she is telling you something that she’s not supposed to repeat. Your voices are lower, and you make sure no one is listening. You are old friends and she trust you to not repeat it further.; the conversation and now shared secret helps strengthen your sense of loyalty, of having a bond that comes from trusting and being trusted.

Like all our experiences, that secret-sharing sensation of trustfulness occurs within the brain. While many calculations may have contributed to your friend’s decision to confide that secret (she wanted your advice or opinion, she thought you’d be interested and it would be a conversational gift, etc.), anticipation of the feeling of closeness the conversation engendered may well have played a part also. Neuroscience research is now discovering the neural correlates of these social tendencies. With trust, a key element is oxytocin.

Oxytocin is a neuropeptide that, among its many functions, plays a big role in facilitating trust. It is an ancient chemical, found in many species, that supports parent-child attachment and strengthens the bonds of mated pairs. Among humans, it also influences social interactions by rewarding prosocial behavior (Hung, Neuner et al. 2017); it released during sex, nursing, touching, and friendly social interaction<sup>4</sup>. It increases

---

4. Including conversations between mothers and daughters Seltzer, L. J., T. E. Ziegler,

people's willingness to trust others (Kosfeld, Heinrichs et al. 2005) and decreases their fear of being betrayed (Baumgartner, Heinrichs et al. 2008)—including when sharing secrets (Mikolajczak, Pinon et al. 2010). It encourages us not only to trust, but also to be trustworthy (Zak, Kurzban et al. 2005). Though our understanding of the neurochemical correlates of behavior is far from complete, it is clear from these and other studies that successful interactions that involve trust are intrinsically rewarding, beyond whatever social or financial benefit the interaction may have yielded: to be trustworthy, and to trust someone and have that trust honored creates a positive emotional experience, reinforcing that relationship and encouraging future trusting behavior.

What would happen if we did not have these intrinsic emotional rewards for trust? The answer shows up in the newspapers every year or so, in a story about a charming but, it turns out, utterly untrustworthy man or woman who repeatedly cultivates friendships, solicitously earning the trust of others only to betray it in a blizzard of unpaid loans, forged checks, and siphoned banks accounts.

Psychopaths are people who lack or have very limited social emotions, including the emotions associated with affective trust. Lacking empathy, they do not feel the pain of others; lacking guilt, they do not worry about the pain they cause. It is estimated that about 2% of the population can be classified as psychopathic (not to be confused with “psychotic”). Some live ordinary lives (indeed, they are highly represented in sales and entrepreneurial careers), but others become criminals, often successfully, for they are unhampered by guilt for their victims—or by any desire to be trustworthy. As I write this, a headline new report features the story of Tracii Hutson, whose warm personality let her make close friends quickly and easily—whom she then cheated out of hundreds of thousands of dollars. Her victims describe her as wonderfully friendly, even charismatic; some repeatedly forgave her, only to be ripped off once again. Her arrest record shows this pattern has gone on for years.

Russell Hardin often uses as an example a story from Dostoyevsky's *The Brothers*

---

et al. (2010). "Social vocalizations can release oxytocin in humans." Proceedings of the Royal Society B: Biological Sciences **277**(1694): 2661-2666., social mirroring Spengler, F. B., D. Scheele, et al. (2017). "Oxytocin facilitates reciprocity in social communication." Social Cognitive and Affective Neuroscience **12**(8): 1325-1333., and gazing into one's dog's eyes Nagasawa, M., S. Mitsui, et al. (2015). "Oxytocin-gaze positive loop and the coevolution of human-dog bonds." Science **348**(6232): 333-336..

Karamazov. A lieutenant colonial engages in a series of secret investments with a local merchant: he lends the merchant money for a short period, and in return shares in the profits. Because it is irregular, to say the least, they handle this investment in secret. Then the lieutenant colonial is given notice that he is being transferred. He goes to the merchant to pick up the money and the most recent “present”, but the merchant pretends to have no idea what he is talking about, a ruse that lets him pocket the full amount of the last installment. It is a cold and calculating move, one that we certainly would not expect if the two men had developed a truly trust-based relationship, with its attendant emotions. It does pass Hardin’s definition of a trust relationship—indeed, it is his featured story— but I would argue instead that the merchant has the emotion-less profit-maximizing response that seems possibly psychopathic, suspiciously like *Homo Economicus*, and certainly not trust-promoting. The lieutenant colonial, like Hutson’s erstwhile friends, put his trust in one who was not worthy, and then experienced the much more negative emotions of trust betrayed.

For the psychopath, their lack of social emotions can be adaptive and profitable. Though they are not trustworthy, they are skilled at eliciting trust and otherwise deceiving those around them that they are well-intentioned; one study showed psychopaths were rated as more trustworthy than neurotypical subjects. The destruction they can wreak is a vivid reminder of the importance of the affective component of genuine trust and trustworthiness.

[the next chapter picks up on this thread to talk about artificial agents and their ability to convincingly display signals of emotion (and trustworthiness) without actually possessing either.]

As a teenager, I sometimes hitch-hiked with friends. After a long stretch of cars whizzing past us on the highway, one would finally pull-over for us. We’d have a few seconds to evaluate it, waving it away if it was too broken down, filled with too many guys or driven by too off-putting a driver. Otherwise, we piled in and hoped for the best. I never hitched alone, and even with a friend or two, found some of the rides we had—jittery speeding drivers, drivers who joked about kidnapping up—too white-knuckled anxiety provoking for my taste.

Today, I do not hesitate to hop into a stranger’s car, alone, day or night. The difference is that my safety (and that of the driver, too) is ensured by the surveillance of a ride-hailing app. The company knows who I am, who the driver is, it knows where we are supposed to be going and where we are now. Yes, there have been occasional stories of assaults by Uber drivers but for the most part, it is a quite safe arrangement. Especially when you consider that in certain places, driving a taxi is one of the, if not the number one, most dangerous profession. (Indeed, Gambetta made a close sociological analysis of the rules and heuristics taxi drivers in especially tense and dangerous areas used to determine which passengers

they would take: too selective and they would not make any money; not selective enough and they would be at risk of beatings or robbery). Unlike the intense scrutiny passengers and drivers in un-surveilled cars cast on their potential companions in the confined and mobile car, ride-hailing passengers are often dangerously oblivious—enough so that the companies repeatedly remind customers to check the license plate of the car — not a few have been so inattentive that they climbed into a random car. Otherwise, no evaluation needed, or really allowed: the driver is not supposed to turn down passengers and vice versa.

Though I found hitch-hiking in the congested anonymous built-up suburbs of my teens to be nerve-wracking, many writers, have seen in hitching a heady combination of trust, risk and discovery; Jack Kerouac's "On the Road" may be the primary example. Hitchhiking was common in the first half of the 20th century it was common. One reason is that cars were rare. But it was arguable a more trusting time, and giving and requesting rides was, in the words of one transportation historian, a collective good, a "'we're all in this together' sort of thing" (Pfleger 2015 )

Today, hitchhiking thrives in a few places, but is a thing of the past in most of the world. Yet some people feel passionately about it, for the opportunity it provides to explore trust and vulnerability on the open road. A couple who have been hitch hiking around the world wrote: "'In both relationship or hitchhiking, we have the risk of something bad happening to us but at the end of the day, it's our choice to trust or not, to open ourselves to amazing possibilities or to limit ourselves because of fear".<sup>5</sup> Josh DeLacy, who hitched across the country in 2013 and kept a popular blog about his travels, called it "Traveling on Trust"; a repeated theme in it is about people feeling as if they are repaying the favors that had been done for them when they were younger and traveling—trust repaid across time and space. These and other accounts all mention the experience of discovering trust with a stranger as one of the most meaningful aspects of it.

Hitchhikers talk with their drivers. Passengers in ride-hailing apps, not so much. Most, the drivers say, are on their phone. A few, the most chatty or curious, do talk, but it is typically a superficial chatter, the conversation of people who have no need to interact and will part ways shortly. Yet even that minimal connection will likely be gone in the future, when autonomous vehicles replace much of the fleet. How will we decide whether or not to have confidence in them?

---

5. <https://maptheunknown.com/2019/03/17/is-hitchhiking-safe-the-other-side-of-trusting-a-stranger/>



The rise of ride-hailing apps is but one of many examples of constraints and surveillance replacing trust as the means of regulating our interactions. AirBnb has played a similar role in the world of house sharing/renting. Many of our conversations take place in monitored and moderated online spaces. Algorithms determine who receives a loan, a reprieve, a reward. Yet trust is valuable, both for the relationships it stewards and as an end in itself. As our world becomes bigger and our interactions move online, trust alone cannot suffice, but often it can be supplemented, rather than replaced. For this to happen, though, we need to consciously value trust, and be willing to take on the added social effort it takes to establish and maintain trusted relationships.

That said, we also need to be cognizant of the insular and sectarian aspect of trust. We trust those who are like ourselves. A tightly bonded group of people with strong ties to each other has a high level of trust among the members of that group. It is insular—difficult or impossible for outsiders to join, and the criteria for joining may be based on discriminatory factors—race, religion, gender. If being in that trusted group provides access to resources, power, such discrimination adds to the entrenched biases that much of contemporary society is endeavoring to reduce.

Such groups play their most important role when the outside world seems hostile, and people want to interact only with trusted others. The trusted group bands together to carry out actions that are forbidden by the larger world. Such a group needs especially strong bonds of trust, for they can otherwise be betrayed to the outside authorities. Their continuing activities create greater trust among themselves, as they learn more about each other—and the more information they have on each other—the more tightly they are bonded to each other. (Oxytocin also supports the parochialism of trust, making people significantly more likely to trust and support members of their own group, but not outsiders (De Dreu, Greer et al. 2011; De Dreu 2012)).

The morality of such groups depends on what the surrounding culture is, and why they are operating illicitly within in. Many examples are of various types of criminal enterprises, whether Mafia families, Wall Street traders using illegal strategies, street gangs. Operating against the laws of the surrounding community, they work to enrich themselves at the expense of others. Yet the same structures, the same relationship to the larger culture is true of moral groups opposing an encompassing and immoral society, such as the partisan groups that resisted the Nazis during World War II or, the Underground Railroad in the United States before slavery was abolished. Consider a small group of prisoners plotting to escape—they require an extremely high level of trust among themselves—but the morality of their endeavor depends on whether we think they were justly or unjustly imprisoned, whether we think the government that imprisoned them is a moral authority.

## References

- Baumgartner, T., M. Heinrichs, et al. (2008). "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans." *Neuron* **58**(4): 639-650.
- De Dreu, C. K. (2012). "Oxytocin modulates cooperation within and competition between groups: an

- integrative review and research agenda." Hormones and behavior **61**(3): 419-428.
- De Dreu, C. K. W., L. L. Greer, et al. (2011). "Oxytocin promotes human ethnocentrism." Proceedings of the National Academy of Sciences **108**(4): 1262-1266.
- Hung, L. W., S. Neuner, et al. (2017). "Gating of social reward by oxytocin in the ventral tegmental area." Science **357**(6358): 1406-1411.
- Kosfeld, M., M. Heinrichs, et al. (2005). "Oxytocin increases trust in humans." Nature **435**(7042): 673-676.
- McKnight, D. H. and N. L. Chervany (1996). "The meanings of trust."
- Mikolajczak, M., N. Pinon, et al. (2010). "Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance." Biological psychology **85**(1): 182-184.
- Nagasawa, M., S. Mitsui, et al. (2015). "Oxytocin-gaze positive loop and the coevolution of human-dog bonds." Science **348**(6232): 333-336.
- Pfleger, P. (2015 ) "Thumbs Up, Then And Now: Hitchhiking Stories From The Road." NPR.org.
- Putnam, R. D. (1995). "Bowling alone: America's declining social capital." Journal of democracy **6**: 65-78.
- Rainie, L., S. Keeter, et al. (2019). Trust and Distrust in America. US Politics and Policy, Pew Research Center.
- Seltzer, L. J., T. E. Ziegler, et al. (2010). "Social vocalizations can release oxytocin in humans." Proceedings of the Royal Society B: Biological Sciences **277**(1694): 2661-2666.
- Spengler, F. B., D. Scheele, et al. (2017). "Oxytocin facilitates reciprocity in social communication." Social Cognitive and Affective Neuroscience **12**(8): 1325-1333.
- Xin, Z. and S. Xin (2017). "Marketization process predicts trust decline in China." Journal of Economic Psychology **62**: 120-129.
- Zak, P. J., R. Kurzban, et al. (2005). "Oxytocin is associated with human trustworthiness." Hormones and behavior **48**(5): 522-527.