Online Trust, Trustworthiness, or Assurance?

Coye Cheshire

Abstract: Every day, individuals around the world retrieve, share, and exchange information on the Internet. We interact online to share personal information, find answers to questions, make financial transactions, play social games, and maintain professional and personal relationships. Sometimes our online interactions take place between two or more humans. In other cases, we rely on computers to manage information on our behalf. In each scenario, risk and uncertainty are essential for determining possible actions and outcomes. This essay highlights common deficiencies in our understanding of key concepts such as trust, trustworthiness, cooperation, and assurance in online environments. Empirical evidence from experimental work in computer-mediated environments underscores the promises and perils of overreliance on security and assurance structures as replacements for interpersonal trust. These conceptual distinctions are critical because the future shape of the Internet will depend on whether we build assurance structures to limit and control ambiguity or allow trust to emerge in the presence of risk and uncertainty.

The Internet is an extraordinary tool for human interaction, connecting millions of people with various information technologies and systems across a global communication network. Despite the advantages of the Internet for communication breadth and efficiency compared to offline interactions, there are legitimate concerns about what it means to trust others in a shared Internet commons. For example, the social cues we rely on to detect risk and uncertainty in the physical world are often unreliable when we do not know who is behind the digital curtain of anonymity. We interact online to share personal information, find answers to questions, make financial transactions, play social games, and maintain professional and personal relationships. Sometimes our online interactions take place between two or more humans using different tools and services, including social networking websites, chat clients, messaging sys-
tems, blogs, and forums. In other cases, individuals are not directly involved in information transfers; we allow computer systems to act as agents, managing and exchanging our information according to specific rules and heuristics.

In each type of interaction mentioned above, risks and uncertainties complicate our decisions. Should we post pictures from a recent party to a social networking site? Can we believe the highly negative evaluation of a local restaurant on a review site? Are our medical search histories safe with our favored search engine companies? In some cases, the stakes for imparting our trust may be relatively low, such as when we evaluate the validity of an email chain letter or choose a restaurant based on a few anonymous reviews. However, when real financial and material assets are involved and there is ambiguity about a given outcome, trust is imperative. Valuable assets constitute risk (or what is at stake) in an interaction; the numerous sources of ambiguity about an outcome create uncertainty.

The combination of risk and uncertainty is critical for understanding trust in any situation, whether online or offline. Some situations might carry high risks, such as financial assets or personal reputation, but little or no uncertainty about the outcome. In other circumstances, nothing of significant value is at stake, but the outcome is indeterminate. I agree with those who argue that trust is relevant only when risk and uncertainty exist together.\(^2\) In this view, trust is not simply the absence of risk and uncertainty. More accurately, trust is a complex human response to situations that are rife with risk and uncertainty.

Given the many risks and sources of uncertainty that exist in online interactions, it is unsurprising that trust remains a principal topic in social and technical studies of the Internet. In addition to the interpersonal communications that occur between humans on the Internet, the Internet infrastructure itself (computers, protocols, and connections) depends on reliable and secure relationships between systems. The Internet is a complex arrangement of layered trust relationships, including small networks among hardware systems, protocols that often presume reliable relationships between these networks, human administrators who configure and maintain the disparate networks, and, finally, the end users who operate personal computers and devices to interact with other humans and systems.

This essay examines key concepts such as trust, trustworthiness, cooperation, and assurance in online environments. The distinction between interpersonal trust (human to human) versus system trust (human to system) is particularly important as we begin to think about the future of trust on the Internet. To many social scientists, it makes little or no sense to treat systems as autonomous entities in the same way we regard humans. Likewise, some computer scientists who worry about the security and reliability of systems may see little in common between the uptime or security of an Internet server and the ambiguities of friendships and online social relationships. However, there is much to gain from considering different meanings of trust in the context of specific risks and various sources of uncertainty. The interests of those who endeavor to build a more reliable and secure Internet may not be so different from those who use the Internet for countless social purposes. In each case, the concept of trust can help define and describe the complexity of anticipating imminent outcomes and behaviors in the presence of uncertainty.

Some scholars argue that the term trust is accurate only when applied to interper-
sonal, human relationships. According to this view, interpersonal trust develops over time in a direct relationship when one party believes the other party has incentive to act in her interest or take her interests to heart. Eventually, an enduring relationship can develop, consisting of either indirect or direct interactions. The continually fulfilled expectation of trustworthy behavior from another person over time defines a trust relationship.

Intuitively, we separate the deeper sense of trust that we evoke when we speak of family, friends, and other individuals with whom we share close ties from the term’s more mundane usages (as in, for example, the statement “I trust the mail carrier to pick up on time”). The same issue extends to online environments. On popular online social networking services, such as Facebook, we might not trust every friend in the same way or in the same circumstances. One way to disentangle the complexity of interpersonal trust is to consider carefully what constitutes a trusting relationship in a specific context. Trust is a useful and meaningful term when we consider the possibility that not all friendships and acquaintances are, in fact, trust relations.

Political scientist Russell Hardin describes significant trust relationships as those in which each party encapsulates the other’s interests. For example, consider two workers who collaborate on a complex project. If the individuals trust one another, then they each believe the other is trustworthy enough to perform a certain type of task in a competent way. One person may entrust her work for modification or enhancement by the other and continue to do so in a reciprocal fashion over time. In Hardin’s encapsulated-interest framework, the project is important and valuable to both individuals, and each understands the value to the other. This example also emphasizes the fact that trust is largely dependent on the specific situation of social interaction. We might trust various people in different contexts, but trust very few (if any) people in every context.

On the Internet, it may be difficult or impossible to develop encapsulated interest between individuals with ephemeral communications and very little at stake (for example, public chat rooms or threaded conversations on a range of topics). However, in popular forms of Internet communication such as online dating, trust is intrinsic to every aspect of the user experience. Beyond its popularity as a tool for finding romantic relationships, online dating provides a unique window into the complexities of building online interpersonal trust over time through different forms of communication. Online dating is largely about learning to use the affordances of online communication channels with low personal risk, with the purpose of finding individuals who are, among other things, sufficiently trustworthy to meet in person.

Online dating interactions typically follow trajectories that begin with highly uncertain, but low-risk, online interactions with others (such as online messaging and inquiries about possible interest). Interactions are initially uncertain because individuals know only the version of each other presented in their dating profiles. Individuals have a high degree of control over how much they want to risk revealing about themselves in different channels of communication (including instant messages, email messages, online chats, telephone conversations, and face-to-face interactions). If the process of learning about a potential date is successful, individuals may continue their communication offline, where physical and personal risks are arguably more numerous.
Although trust and trustworthiness appear interchangeably in common vernacular, they are distinct concepts. Trustworthiness is a characteristic or property of an individual; trust is an attitude or belief we have about those who are trustworthy. Unlike interpersonal trust, an assessment of trustworthiness does not require prior firsthand communication or experience.

A common way to infer the trustworthiness of another in online environments is through explicit or implicit third-party reputation information. If we receive online advice from someone we do not already know, even basic information about others’ experiences with the individual can help us make an informed decision about the credibility of his or her claims. Explicit reputation information includes ratings, reviews, and other assessments of his or her qualities based on prior personal experience. Alternatively, implicit reputation information includes signals, cues, and traces from an individual’s prior actions that might correlate with future behavior (for example, his or her frequency of activity in an online forum or accuracy of grammar and spelling). Online reputation information is not a perfect solution to the problem of trust because it is vulnerable to exploitation, deception, and misinterpretation. Despite these challenges, stable reputation systems can ameliorate concerns about risk and uncertainty, leading to online cooperation and higher assessments of trustworthiness.

Many different issues can affect judgments of trustworthiness, including the nature of the situation and perceptions about another person’s intentions and motivations. When we assess an individual’s trustworthiness, we essentially decide whether to take a risk with that person. Typically inconsequential features and characteristics become essential for inferring trustworthiness, even if our assumptions about the link between certain characteristics and behavior are imperfect. This observation extends to the online world of social interaction, where email addresses, domain names, and other features imply trustworthiness in the absence of other available attributes.

Given that relational trust requires ongoing, experiential information about another person, nonrepeated interactions between individuals with no prior communication are not based on trust: they are acts of risk-taking. In interpersonal online interactions, an act of risk-taking does not guarantee reciprocity or the development of trust. In fact, experimental research in computer-mediated environments demonstrates that risk-taking among anonymous actors often goes unreciprocated. Numerous online interactions are fleeting, one-time communications. Examples include question-and-answer forums, threaded comments on websites, blog commentaries, and public online chat rooms. In these situations, assessments of trustworthiness can be essential, even if relational trust is not possible. For this reason, to call for more trust in anonymous online interactions is misleading. Relational trust is not even possible or desired in these situations, but the ability to infer trustworthiness is essential.

Signaling one’s intentions through an initial act of risk-taking is critical in online interactions when incomplete information (anonymity or pseudonymity) makes it difficult to assess the trustworthiness of others. Risk-taking can act as a signal when individuals intentionally give up something of value without any explicit form of assurance. For example, taking a disproportionate amount of risk in an online transaction by sending money before an item has shipped sends a clear signal that the buyer believes the seller is trustworthy. Willful acts of on-
line risk-taking can also send a more general signal about an individual. As media and social communications scholar Judith Donath argues, “[R]isk taking is another behavior that may seem irrational, but when viewed as a signal, can be seen as a way of claiming a high level of fitness…. [P]osting revealing or culpable material online arguably has become another forum for signaling imperviousness to danger and repercussions.” Those who show that they are willing to take a chance online open themselves to risk; however, the long-term payoff is a kind of online social intelligence that rewards risk-taking with new opportunities.

If individuals cooperate with one another over time, is it accurate to say that the individuals are engaged in a trusting relationship? There are roughly two schools of thought on the link between trust and cooperation, but both perspectives agree that the two concepts are separate. Scholars such as political scientist Robert Putnam argue that trust is required to produce cooperation, which in turn helps create productive societies. A competing viewpoint maintains that trust exists at the interpersonal level to produce “social order and to lower the costs of monitoring and sanctioning that might be required if individuals were not trustworthy.” In this latter view, cooperation can (and often does) exist independently of trust.

Conceptualizing the difference between trust and cooperation requires an understanding of motivations and the context in which individuals interact. For example, elected officials on different sides of the political spectrum may cooperate on legislation to further their own (perhaps competing) interests. Similarly, individuals who agree to exchange valued goods and services through popular online classified advertisement websites (such as Craigslist) may cooperate by successfully completing an agreement or exchange. In both examples, cooperation can exist without requiring interpersonal trust. However, “trust is most likely to emerge in contexts in which the parties find themselves in ongoing relationships,” argue sociologist Karen Cook and her colleagues. Trust may not be required in cooperative interactions but becomes increasingly important as relationships persist.

Experimental research clearly demonstrates that acts of trust are distinct from acts of cooperation. In one laboratory study of social exchange in a computer-mediated system, my colleagues and I measured entrusting behavior as the number of valued items a participant entrusted to another. Individuals could keep the valued items or return them to the entruster (who would then receive double the entrusted value in return, paid by the experimenter). Cooperation, therefore, was the decision to return rather than keep the entrusted items. The results of the study show that individuals entrust very little to their partners when they exchange with a new, random partner on every interaction opportunity. However, when individuals interact with the same cooperative partner over time, they tend to increase slowly the amount entrusted. Thus, cooperation over time leads to larger, higher-risk entrustments. The occurrence of ongoing risk-taking in the presence of uncertainty makes trust both possible and relevant. It is the process of risk-taking with the same individual over time that separates one-time assessments of trustworthiness from interpersonal trust.

When organizational or institutional mechanisms exist to protect individuals from harm (such as betrayal), individuals tend to rely primarily on the third-party protections instead of building mutual
trust. This inclination leads to a trust paradox because the assurance structures designed to make interpersonal trust possible in uncertain environments undermine the need for trust in the first place. Individuals have a strong incentive to act cooperatively when robust monitoring and assurance structures are present, but cooperation in these cases has more to do with sanctions and other negative outcomes than interpersonal trust. In essence, strong forms of online security and assurance can supplant, rather than enhance, trust. A different way of viewing the trade-off between trust and assurance requires returning to the preconditions of risk and uncertainty that underlie interpersonal trust relations. Effective assurance systems minimize or eliminate one or more forms of uncertainty, such as the ambiguity of another’s intention to follow through with a prior agreement. The source of uncertainty then shifts to the assurance system, thereby making trustworthiness and reliability of the institution or organization the salient relationship.

Strong assurance systems are widely portrayed as a solution to the problem of trust in Internet environments. For example, third-party verification systems for buyers and sellers who use Internet auctions help reduce malfeasance and fraud. In addition, a leading suggestion for addressing the reliability of routing information on the Internet involves the concept of a trust anchor, or an authoritative body that provides assurances about data authenticity between Internet networks. However, my research with Ashwin Mathew, Ph.D. candidate at the University of California, Berkeley, demonstrates that the complexities of Internet routing are currently resolved through interpersonal trust relationships among network administrators. Ongoing trust relationships enable the network administration community to act collectively and dynamically to keep information flowing properly through the disparate networks that form the larger Internet. Replacing human interpersonal relationships with authoritative trust anchors would provide a centralized point of decision-making, but also of potential failure.

The choice between assurance structures or emergent trust relationships without structural assurances on the Internet is largely about short-term versus long-term goals. Assurances lessen uncertainties, while trust thrives where uncertainties are abundant. Although individuals cooperate as if they trust one another when third-party assurance structures guarantee interactions, this behavior is ultimately unstable. When and if institutional assurance structures fall short, individuals realize that the original source of security is gone, and interpersonal relationships must be forged once again amid uncertainty.

Experimental research on computer-mediated trust relationships supports the claim that assurances may solve short-term problems of cooperation at the expense of building long-term trust. In laboratory research on trust and forms of social exchange, my colleagues and I find that individuals build high levels of trust when partners cooperate in highly uncertain environments, compared to those who cooperate in low-uncertainty environments (namely, when interactions are ensured by a third party). However, when individuals who have built trust with their partners without assurances shift into a new interaction environment with assurances, trust significantly decreases between the pair even when cooperation remains high. As individuals learn to rely on assurance structures, they do so at the expense of interpersonal trust. The key implication is that if assurances fail, cooperation and trust will fail as well.
My discussion of trust and related concepts has thus far been limited to interpersonal, human-to-human interactions. However, many researchers and practitioners primarily focus on the relationship between humans and the technologies, systems, and interfaces that we use on the Internet. Online trust has been the focus of a wide variety of research in Internet studies, computer-mediated communication, human-computer interaction, computer-supported cooperative work, and related fields. Among the more technical areas in Internet research, the term trust routinely refers to several related but distinct concepts, including credibility, security, surety, and reliability. Unfortunately, these numerous meanings create an overabundance of complex models that are largely incompatible because of the vast conceptual differences among key terms.

Trust in an information system primarily involves individuals’ expectations about whether the system will operate in a predictable manner and provide reliable outputs. While many may feel that a hard-line distinction between human-to-human and human-to-system trust is unnecessary, the potential difference in meaning for terms such as intention, agency, cooperation, and choice is difficult to ignore when considering programmed systems versus human actors. For example, philosopher Annette Baier argues that relational, interpersonal trust depends on the possibility of betrayal by another person. Information systems and computer programs appear to lack the agency and consciousness to choose freely to betray the trust that users place in them.

In many situations, there may be little perceived difference in the evaluation of risk and uncertainty when interacting with a system or with another person. From the perspective of humans who believe that they trust systems, evaluating the trustworthiness of a system is very much like assessing the trustworthiness of a person because an individual will use his or her prior experiences with the system (and other similar systems) to create an estimate of risk and uncertainty. However, the interaction lacks the relational dynamic that is essential to interpersonal trust. Even if an individual develops very complex heuristics for “trusting” systems, it is often a unidirectional evaluation.

Unlike complex social relationships in which humans must continually assess one another’s motivations, intentions, and behaviors before and after an interaction, it is difficult (and currently implausible) for an online system to be endowed with the sentience and free will to shift between different motivations and intentions unless these are programmed, defined, or changed by other humans. As computer scientist John Seely Brown and historian and social theorist Paul Duguid demonstrate through numerous examples, people tend to treat bots and information agents as if they are human, even when we clearly know they are not.

In addition, the experimental work of communications scholar Cliff Nass and his colleagues provides further empirical evidence that people anthropomorphize computers and programmed systems that respond like humans.

It is also possible that “trust-like” behavior in computers and systems is basic categorization and simplification of the sort that linguist George Lakoff describes as fundamental to human understanding. According to Lakoff, individuals lump similar concepts together based on personal experiences, a process that is often good enough for routine sense making. We might feel betrayed by a computer even if we fully comprehend that the computer did not really choose to
act a particular way. Thus, equating trust in a computer system with trust in another person can seem completely reasonable from an individual’s perspective.

Distinguishing between interpersonal trust and human-system “trust” may seem somewhat pedantic, especially outside of scholarly debate. However, when we delve deeper into the realm of building enduring trust relationships and constructing trustworthy networks and systems, the distinctions between concepts like interpersonal trust and system trust become essential for understanding when and why trust fails. When a human betrays a friend’s trust, the friend knows who is culpable, and the consequences are often clear for both parties. When a system “betrays” a human’s trust, assigning blame can have enormous repercussions: should we blame the system itself, those who programmed the system, the organization that hosts the service, a quality control person, the director of the organization, or the organization in general? Can the system learn or do anything differently after failed trust? Put simply, we need to be able to attribute an outcome to someone or something if we hope to understand and respond to failures of trust in systems.

In late January 2011, Egyptian authorities successfully shut down the international Internet access points that allow traffic to pass to and from systems in Egypt. Shutdowns of this type occur from time to time for technical and political reasons, but this event was largely unprecedented in terms of range and scale. The explanation for the Internet connectivity blackout in Egypt emerged from a straightforward political decision: the autocratic Egyptian government hoped to quell information sharing and coordination among its citizens amid the growing talks of protests against the current regime.

The Internet blackout in Egypt was not really about system trust, nor was it about direct, interpersonal trust between individual citizens and government officials. It ultimately concerned the Egyptian government’s lack of trustworthiness in the eyes of the Egyptian people as a result of various actions over time. The government’s decision to interrupt Internet access furthered the image of an untrustworthy regime, leading to contempt that spurred calls for revolution.

The event in Egypt is an example of what sociologist Niklas Luhmann views as a type of system trust in governments and other organizational machinations. In Luhmann’s macro-level view of system trust, individuals trust the structures of human interaction and organization. However, I believe that the Internet shutdown in Egypt is a model example of why trust placed in information systems often has less to do with the actual systems and more to do with the humans, organizations, and governments that maintain or control them. In contrast to Luhmann’s position, Russell Hardin argues that trust in macro social structures (governments and other human systems) is largely implausible. Hardin believes that the complexity and scope of any large government makes true interpersonal trust impossible; indeed, each citizen cannot realistically build firsthand experience with every politician or official. Thus, the relational view of trust maintains that it is more important “that government be trustworthy than that it be trusted.”

Hardin’s position is consistent with Luhmann’s opinion if we view each as a statement about trustworthiness rather than trust.

Creating and maintaining social order and trust between individuals are fundamental problems for human interaction, whether online or offline. Many existing and emerging online systems enable
social interaction, collective action, and interpersonal communication on a scale that was unthinkable before Internet use became commonplace. The Internet is the real world, but we must remain cognizant of the complexities of trust and social interaction in computer-mediated environments. Our ability to understand and articulate the differences between trust, trustworthiness, cooperation, assurance, and related concepts is not a purely bookish problem: the possibilities and limitations of trust and social interaction on the Internet will depend entirely on how we design online communication technologies in the context of the surrounding global political and institutional environment. Will we continue to facilitate interpersonal interaction and embrace the complexity of emergent social uses of online information technologies? Alternatively, will we build structures to control, limit, and secure online social interactions—and accept the potential trade-offs for trust in exchange for assurance?

Clarifying the differences between forms of online trust is crucial because current and future policy-makers justify strategies and decisions with outcomes and recommendations from trust and security research. In spite of attempts to create trust by eliminating doubt and minimizing peril, there is no quick way to build meaningful trust without overcoming real risk in the presence of uncertainty. Assurance structures are undoubtedly an important part of complex systems; sufficient evidence shows that they can encourage cooperative behavior—especially when all other options fail. However, if we attempt to stamp out all online uncertainty and risk through security measures and centralized assurance structures, we may inadvertently create a less trusting Internet environment in the long term.

ENDNOTES

1 I am deeply grateful to Ashwin Mathew for his insightful comments and suggestions on an earlier draft of this essay.

2 Russell Hardin, Trust and Trustworthiness (New York: Russell Sage, 2002).


5 Andrew Fiore and Coye Cheshire, “The Role of Trust in Online Relationship Formation,” in Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives, ed. Dominika Latusek and Alexandra Gerbasi (Hershey, Pa.: IGI Global, 2010), 55 – 70.


Online Trust, Trustworthiness, or Assurance?


10 Coye Cheshire and Judd Antin, “None of us is as lazy as all of us: Social Intelligence and Loafing in Information Pools,” Information, Communication & Society 13 (4) (2010): 537 – 555.


13 Ibid., 4.


27 Hardin, Trust and Trustworthiness, 152.