From "Trust and Distrust in Organizations:
Dilemmas and Approaches."

Roderick M. Kramer & Karen S. Cook, Editors

Volume VII in the Russell Sage Foundation
Series on Trust.

Russell Sage Foundation, New York, 2004

Chapter 7

Will Security Enhance Trust Online, or Supplant It?

HELEN NISSENBAUM

PROMOTERS OF the Internet and other digital media cite many and diverse benefits of these advances to humanity, from wide-ranging access to information and communication to enhancement of community and politics to stimulation of commerce and scientific collaboration. As the digital infrastructure has grown in size and complexity, however, even the most enthusiastic proponents acknowledge that the benefits are not inevitable but rest on a number of contingencies. Key among them is trust. Just as in conventional settings where trust improves the lives, prospects, and prosperity of individuals, relationships, and communities, so would it online, and just as distrust can degrade many dimensions of life, so can it online.[1] Because of its importance to a flourishing online world and because trust online poses distinctive challenges, it has attracted the attention of a number of scholars, researchers, and practitioners in technical as well as nontechnical fields, which has yielded an extensive literature in scholarly and trade publications, the popular media, and government reports, and an active showing at conferences.[2]

Much of this work on trust online is devoted either to empirical investigation of key variables relevant to trust online or to developing technical models, mechanisms, and designs for encouraging and supporting trust online. While acknowledging the practical contributions of both these lines of work, this paper questions the conceptual assumptions behind them, arguing that the successful end-point of these efforts will not be trust at all but something else, something closer to surety, or the certainty that one is secure, particularly through technically imposed constraints.

Background

The impetus for studying trust online has emerged primarily from two sources: concerns of technical experts over computer and network security, and concerns of proponents of ecommerce over possible pitfalls of the new medium. By the late 1990s, computer scientists and engineers had come to discuss goals of the technical field of computer and network security in terms of trust, seeking to build "trusted" or, rather, trustworthy, systems.[3] Trust came to be viewed as an aggregate or higher-order property of systems manifesting a constellation of other valued properties, including integrity, availability, survivability, and so on (Schneider 1999). These experts worried that our vast networked information system—the network of networks that includes local private systems as well as public systems like the Internet, the Web, cyberspace—is vulnerable to technical failure as well as malicious attack.[4] Those whose interest in trust stemmed from an interest in ecommerce understood that consumers would balk if they were fearful of being harmed in any of a variety of ways, such as being defrauded, having their credit card numbers stolen, or receiving poor-quality goods. Businesses, too, would stay away if they anticipated costly losses from failure to pay, repudiation of customer commitments, technical failures, and so on.[5]

Although the sources of concern over trust are distinct, there is significant overlap in the nature of the solution envisioned—namely, a suite of technical security mechanisms aimed at inducing users in various roles to trust networked information systems and one another. Mechanisms that would ensure trustworthy systems would in turn induce consumers to trust providers of goods and services; providers, to trust consumers; and both of these, to trust the vendors of underlying technical systems and in general would engender a climate of trust online.[6] So conspicuous has been the vision of trust through security portrayed by these two groups that it currently occupies the mainstream—in part because there are no equally persistent, competing interpretations, and in part because talk of trust online is relatively new and in the mainstream view, relatively uncontested. Later in this chapter, I shall say more about these mechanisms, but here I would like to label this common vision with a slogan: trustworthiness as security, or *trust through security*.[7]

This chapter articulates a skeptical position on the vision of trust through security; instead I contend that the ideal endpoint of trust through security is surety and not, in fact, trust. This contention draws on a commonsense conception of trust elucidated in a number of important philosophical and other theoretical works. It does not challenge the value of surety itself but argues that striving for surety online is different and sometimes even inimical to striving for trust; conflating the two misconstrues the nature of trust and misses the point of why we care about it. Accepting this claim, however, means that we will need to make an explicit normative (or policy) commitment either to the value of an online environment grounded in surety, or to one grounded in trust.

Conceptual and Technical Scope

Before elaborating on key themes and contrasts, it is necessary to delineate the scope of this paper. The conceptual domain of trust is immensely broad and not always internally consistent; the realm of computing and information and communications media is itself large and varied. We need, therefore, to make certain simplifying assumptions and qualify the scope of our investigation in order to make a meaningful contribution even within this more limited domain.

In its broadest sense, the online world we speak of could cover the entire technological system, vast and powerful, that sits at the hub of almost all other parts of the critical infrastructures of society, controlling—and in some cases conjoining—energy, commerce, finance, transportation, education, communication, and more, and in so doing affecting almost all modes of social, community, cultural, and political life.[8] This essay does not address the system as a whole—the vast and powerful grid that connects and controls satellites, nuclear devices, energy, the stock exchange, and so forth. Instead, it focuses on the parts of the system directly experienced by ordinary people, who in increasing numbers use it to talk, conduct business transactions, work, seek information, play games, and transact with public and private institutions. At present, this realm comprises the World Wide Web (the Web) and the various servers (computers), conjoined networks, people, and institutions that constitute it. It also comprises the realm that at times interacts with the realities of the offline world and at other times fragments into an apparently independent and separate reality that some writers and participants have taken to calling cyberspace, or the "virtual" world.

Neither does this essay cover everything that the word "trust" could mean. Trust is an extraordinarily rich concept covering a variety of relationships, conjoining a variety of objects. One can trust (or distrust) persons, institutions, governments, information, deities, physical things, systems, and more. Here, I am concerned with two ways that "trust" is used. One is as a term describing a relationship between one person (a trustor) and another (the trustee). Although, in practice, the trustee position could be filled by almost anything, here I limit consideration to cases where the trustee is a being to which we are willing to attribute intentions, motivations, interests, or reasons, and might also refer to as "agent." Central in this category are people—individually and in groups; I would also be willing to include organizations, communities,

and institutions. However, in my discussion I exclude from the "trustee" category at least one quite common referent for trust in the online context: the networked, digital information systems themselves, the layered hardware and software that individually constitute the microsystems and the macrosystem that is formed by these. This is not because of any deep-seated disagreement with those who write about trust in relation to networked information systems or information and communications technology and worry about the dependability of these systems, their resilience to various forms of failure and attack, and their capacity to protect the integrity of online interactions and transactions. My reasons are pragmatic. These cases are sufficiently distinct from one another that they deserve separate (but equal) treatment. Following others, I use the term "confidence" to refer to trust in systems, recognizing that trust in the online world begins with confidence in systems, but does not end there.[9]

Conditions of Trust

At the same time that proponents of the Internet acknowledge the key role of trust in enlivening activity, interaction, participation, and institutional growth, they recognize the distinctive challenges to trust building posed by the online realm. To see how these challenges arise, it is useful to consider, first, conditions that have been associated with the formation of trust generally, and then the ways that online mediation affects them; determine mechanisms governing trust, namely factors that systematically affect tendencies to trust (or not to trust) other people, groups, and institutions; and how features of online interaction affect them.

It is worth noting that for the purpose of this discussion, it does not matter whether trust is a species of belief (or expectation)—a cognitive stance—or is a noncognitive attitude, even though this is a matter of some disagreement among theorists and social scientists. Some, such as the philosopher Annette Baier (1986), who assert a version of the former view would probably frame the inquiry into mechanisms in terms of reasons that systematically undergird trust, and may even subject such reasons to judgments of rationality or irrationality.[10] Those, such as the philosopher Lawrence Becker (1996), who defend a noncognitive account of trust would probably frame theirs as an inquiry into factors to which the formation of trust is systematically responsive as cause to effect.[11] Most important, agnosticism on this matter should not block access to empirical and analytic results that link trust with the variety of phenomena that are widely perceived to function as cues, clues, or triggers, whether as reasons or merely causes.

A caveat: The factors listed below should not be understood as a complete account of causes of or reasons for trust formation but should be understood to be selective, reflecting a particular concern for trust in the online context. Furthermore, I acknowledge that my efforts are conceivably incompatible with views on trust—such as those of Adam Seligman (1997)—that reserve the concept of trust for an even more qualified subcategory of attitudes than the one I have articulated above. Seligman would probably say of many of the cases I mention below, where trust is induced by perceived similarity, roles, and other structured relationships, that these are instances of confidence and not trust.[12] To engage further on this point of disagreement—interesting as it is—would deflect us too far from the main subject here. It is important, though, to acknowledge the difference between my more ample and Seligman's more austere concepts. One way to reconcile the difference would be to suggest that followers of Seligman's usage recast the concern of this chapter as being one of trust, faith, confidence, and familiarity online.

History and Reputation

One of the most convincing forms of evidence that others merit trust is their past behavior. If they have behaved well in the past, protected our interests, have not cheated or betrayed us, and in general have acted in a trustworthy manner, they are likely to elicit trust in the future. If they have disappointed us in the past, then we will tend not to trust them. Where we have not built a history of direct interaction with others, we may refer to the experiences of others—we may be influenced by reputations.

Inferences Based on Personal Characteristics

A trusting attitude may be triggered by the presence of perceived qualities in the other. Philip Pettit identifies four: virtue, loyalty, prudence,[13] and a desire for the good opinion of others,[14] all qualities that influence whether a person will trust those who are seen to have them. Pettit writes: "To be loyal or virtuous or even prudent is, in an obvious sense of the term, to be trustworthy. It is to be reliable under trust and to be reliable, in particular, because of possessing a desirable trait" (Pettit 1995, 211). The fourth quality, namely a desire for the good opinion of others, although less deserving of our admiration, is nevertheless a powerful mechanism for preventing betrayals of trust.[15] Accordingly, Pettit recommends against calling the person who chases good reputation trust*worthy*, preferring a more modest commendation of trust-*responsive*, or trust-*reliant*.[16] Though not in direct disagreement with Pettit's characterization, Adam Seligman offers a different perspective, drawing attention to the importance of familiarity, similarity, and shared values as triggers of trusting attitudes.[17] What we know about someone, what we may infer on the basis of "their clothing, behavior, general demeanor," (Seligman 1997, 69) may lead us to judgments about their values and

moral commitments, especially telling if we judge these to be similar to ours. A common religious background, high school, neighborhood, or traumatic experience (for example, having fought in the same war) affects our level of confidence in predicting what others will do and how inclined we are to rely on them. Though related to loyalty, these considerations are not identical. When one depends on a loyal cousin, for example, one counts on the family relationship to induce trust-reliance in one's cousin. Where trust is triggered by familiarity and, perhaps, a perception of shared values, a trustor does not necessarily count on these qualities to cause trustworthy behavior; the trustor merely forms expectations regarding the likely actions of these others.

Relationships: Mutuality and Reciprocity

Aside from personal qualities, the relationship in which one stands to another may bear on the formation of trust. The presence of common ends can stimulate trust. Such cases of mutual ends occur when a person is "in the same boat" as another. When I fly in an airplane, for example, I place trust in the pilot partly because he is in the plane with me and I presume that we have common, or confluent, ends; our fates are entwined for the few hours during which we fly together.

Reciprocity is slightly different, but it, too, can be grounds for trust. In a reciprocal relationship, we trust others not because we have common ends but because each of us holds the fate of others in our hands in a tit-for-tat manner. This may occur, for example, when people are taking turns. The agent whose turn is first deals fairly, reliably, or responsibly with the other because soon the tables will be turned. The relationship of reciprocity admits of great variability. In some cases there is a clear and imminent reversal of roles (this year I am chair of our department, next year you take over); in others it is more generalized (I might donate money to cancer research hoping that when I become ill, these funds will somehow help me). Reciprocity is evident in communities that are blessed with a climate of trust, its members helping those in need and trusting that when they themselves are in need, others will help them.[18]

Role Fulfillment

There is another, perhaps more compelling reason for trusting the pilot of my airplane. After all, the pilot would not trust me, in spite of our common interest in staying alive. Crucial to my trusting the pilot is that he is a pilot, and being a pilot within the framework of a familiar system has a well-articulated meaning. I know what pilots are supposed to do; I am aware of the rigorous training they undergo, the stringent requirements for accreditation, and the status of airlines within a larger social,

political, and legal system. Several of the authors already mentioned have discussed the importance of roles to the formation of trust.[19]

Contextual Factors

One of the most intriguing factors to affect our readiness to trust, beyond those that are tied to what we know about the other, is the nature of the setting in which we act.[20] Such settings can be construed quite locally as families, communities, and towns or can extend to such large and diffuse entities as nations and countries.

Four elements seem relevant. The first is publicity: a setting in which betrayal and fidelity are routinely publicized is likely to be more conducive to trust-reliance, and consequently trust, than a setting in which people can effectively hide their deeds—especially their misdeeds. The second is reward and punishment: settings in which rewards and sanctions follow trustworthiness and betrayal, respectively, are likely to induce trustworthiness and trust. Third, where reward and punishment for fidelity and betrayal are not systematically available, promulgation of norms through other means can effectively shape behaviors and establish a climate of one sort or another. Norms are conveyed through parables, education, local lore, songs, fables, and local appraisal structures. What do these norms convey? Do they condemn betrayal and celebrate fidelity or do they mock gullible marks of confidence tricks and disdain cuckolded spouses while proffering admiration to the perpetrators?[21] Finally, a society can nurture a trusting climate by setting in place, through public policy or other means, various forms of "trust insurance" to provide safety nets for those whose trust is betrayed.[22] A simple example of such a policy is the current arrangement of liability for credit card fraud, which must surely increase people's willingness to engage in credit transactions.

Obstacles to Trust Online

It must be the case that at least some uneasiness comes from the novelty or unfamiliarity of online interaction, which by itself can slow the formation of trust. Citing novelty is not all that informative because even if novelty is a key factor, it is important to investigate why it should produce particular outcomes and whether one particular aspect of the new experience is more material to the outcome than others. We must, therefore, look beyond novelty for those features specific to online interaction that bear on trust. Those that we list below—flexible identity, disembodiment, and inscrutable contexts—cloak aspects of character and personality, the nature of relationships, and settings that normally function as triggers of trust or as reasons for deciding to trust (or distrust).

Flexible Identity

The medium's initial design allowed for agents to obscure identity quite easily, and this status holds, to some extent, into the present.[23] In many online transactions, agents are not compelled to relinquish the identities of their offline selves. Although the capacity for anonymity online is beneficial in a number of ways, it shrinks the range of cues that typically trigger trust. If identity is conceived as a thread upon which interactions with others are strung, then without identity, we lose the capacity to thread together a history of interactions and predict outcomes on the basis of past experiences of either vindicated trust or betrayal. Lacking knowledge of sustained identity also deprives us of a means to learn from the experiences of others whether an agent is trust-reliant, since the construction of reputation is hampered—even if not precluded altogether.

Lacking knowledge of an agent's sustained identity also deprives us of knowledge about the relationships in which we stand to others, for example, whether these are reciprocal or cooperative. Finally, because identity is bound up with accountability, people might presume that anonymous agents are less likely to act responsibly. As a result, people would be less inclined to trust.

Disembodiment

Online there is an opacity not only with respect to others' identities but also with respect to many of the personal characteristics that affect (heighten or diminish) attitudes of trust. We are separated from others in time and space; we lack cues giving evidence of similarity, familiarity, or shared value systems. We may not know the other's gender (male, female, or "other"), age, race, socioeconomic status, occupation, mode of dress, or geographic origins. We lack the bodily signals of face-to-face interaction (see Herz 1995). Are we communicating with a fourteen-year-old girl or a fifty-seven-year-old man posing as a fourteen-year-old girl? Are we selling a priceless painting to an adolescent boy or to a reputable art dealer?[24] Are we sharing a virtual room with an intriguing avatar (in the online world a graphical icon representing a person, frequently in the context of a game or community discussion) or a virtual rapist?[25] We must conduct transactions and depend on others who are separated not only by distance but also by time, who are disembodied in many of the ways in which the nature of their concrete presences typically contributes to our sense of their trustworthiness.

Inscrutable Contexts

The settings of online interactions are frequently inscrutable (sometimes self-consciously so) in ways that affect readiness or inclination to trust.

One casualty is role definition, likely to persist until we develop mechanisms for articulating and supporting social, professional, and other roles. Even with roles that appear equivalent to offline counterparts—for example, "shopkeeper"—we lack explicit frameworks of assurances that support them. Online these institutional expectations are not yet as completely formed. For example, in the typical retail experience of buying an automobile in the United States, the role of "used car salesman" has a quite different connotation for trust formation from that of a salesman representing a high-end vehicle such as Mercedes. As for the roles and terms for them that have emerged in cyberspace (like "sysops," avatars, bulletin board moderators, and so on) that do not have obvious counterparts offline, their duties and responsibilities are even less clearly defined and understood.

Just as roles are still relatively unformulated, so are background constraints and social norms regarding qualities like fidelity, virtue, loyalty, guile, duplicity, and trickery. Are we sure that betrayal will be checked, that safety nets exist to limit the scope of hurts and harms, and so on? Although there is evidence of various groups—social groups, interest groups, cultural groups, hackers—vying to promote their respective norms, the territory remains relatively uncharted, a situation whose complexity is further compounded by the territory's global reach. Participants, especially the majority who are not strongly identified with any one of these groups, can rightly be confused. For them, the most rational stance may be one of caution and reserve.

It is important to note that what I call inscrutability of contexts has a double edge. Many people have observed that it is precisely this quality of cyberspace that is so liberating, enticing, promising. Enthusiasts invite you to participate *because* it is new, different, better, seamless, immediate, unstuffy, truly democratic, and so forth. I am not sure, therefore, that the immediate solution to the problem of inscrutability is a wholesale transfer of existing norms, even if we could bring that about.

Security and Trust

It is not surprising that the task of stimulating trust would fall to computer security experts, security-minded systems managers, and government oversight bodies concerned with computer security. Of course, computer security is not a new concern, but has developed alongside computing itself, responding to changes in the technology and the needs of its rapidly expanding range of applications. Promoters of the new medium see yet another role for security technology. They believe it holds promise for engendering trust online because the very mechanisms developed to fulfill general computer and network security needs also, as a matter of fact, seem to supply some of the elements critical to

trust that are perceived to be missing in the online environment—the missing cues, clues, and triggers that affect the formation of trustful (or distrustful) attitudes and beliefs.

What follows is a brief overview of security mechanisms that have been suggested as ways to achieve a more secure and trustworthy online environment, either by restoring relevant triggers or constraining the possibilities for harm. I have simplified the picture by organizing these mechanisms into three rough categories: (1) access control, (2) transparency of identity, and (3) surveillance. The categories, which largely are my own construction, are an obvious simplification of the broad range of work in computer and network security. My intent is not to describe categories explicitly adopted by computer security experts themselves, nor to suggest that there is a monolithic effort of people and projects, but to provide explanatory clarity relevant to the purposes of discussing trust. The categories reflect functionality, not underlying structural similarities, and, as we shall soon see, are highly interrelated.

Access Control

One of the earliest worries of computer security, from the time when computers were stand-alone calculators and repositories of information, was to guard against unwanted access to the computer and its stored information, to maintain the integrity of the information, and to control distribution of the valuable and limited resource of computational power. Early on, the security mechanisms developed to prevent illegitimate and damaging access involved everything from passwords to locked doors.[26] The demands on computer security mechanisms expanded and became more complicated as networks and interactivity evolved. Vulnerability to intrusion increased because networks opened new means of infiltration—email, file transfer, and remote access—that could not be blocked by locked doors. The infamous Morris worm, which received widespread national attention in 1999, jolted all users into noticing what security experts must certainly have feared: that it was merely a matter of time before vulnerabilities in theory would be exploited in practice.[27]

The Internet, and in particular the Web, has further expanded the modes and extent of interactivity while at the same time exposing participants to new forms of unwanted access and attack. The old fears remain: namely, infiltration by unauthorized persons (hackers, crackers, and so on), damage to information and systems, disruptive software flowing across the Net, information "stolen" as it traverses the networks, terrorists and criminals invading the infrastructure and bringing down critical systems. And new fears emerge: "evil" websites that harm unsuspecting visitors, Web links diverted from intended destinations to others, and disruptive applets—mini applications that visitors to websites can download onto their own systems to enable them to enjoy more

extensive services from that site. Rohit Khare and Adam Rifkin note, "While [you are] doing nothing more serious than surfing to some random Web page, your browser might take the opportunity to download, install, and execute objects and scripts from unknown sources" (Khare and Rifkin 1997, paragraph 4). For example, to view a video clip visitors might need to download a player program in addition to the video files themselves; or to view and interact with financial information provided by a financial services company they may download a mini-spreadsheet program. In the process of downloading the appropriate application, however, the user's computer system is infected with a harmful and often devastating applet. Greater interactivity spells greater vulnerability and a need for more extensive protections. Bruce Schneier, a computer security expert, comments on the almost unavoidable vulnerability of the Internet to attack:

> One problem is the permissive nature of the Internet and the computers attached to it. As long as a program has the ability to do anything on the computer it is running on, malware[28] will be incredibly dangerous.
>
> And anti-virus software can't help much. If a virus can infect 1.2 million computers (one estimate of Melissa infections) in the hours before a fix is released, that's a lot of damage. . . .
>
> It's impossible to push the problem off onto users with "do you trust this message/macro/application" messages. . . . Users can't make good security decisions under ideal conditions; they don't stand a chance against a virus capable of social engineering. . . .
>
> What we're seeing here is the convergence of several problems: the permissiveness of networks, interconnections between applications on modern operating systems, email as a vector to tunnel through network defenses as a means to spread extremely rapidly, and the traditional naïveté of users. Simple patches won't fix this. . . . A large distributed system that communicates at the speed of light is going to have to accept the reality of viral infections at the speed of light. Unless security is designed into the system from the bottom up, we're constantly going to be fighting a holding action (Schneier 1999, paragraphs 1, 6, 9, 11, 12, 13).

Working within the constraints of current network and system architectures, security experts have developed a tool kit of mechanisms to protect people and systems against unwanted and dangerous access. One reason why demands on such a tool kit are considerable is because the agents of unwanted access may be not only people but also bits of code, like applets. Standard techniques like passwords remain in use, fortified where needed by such mechanisms as "firewalls," which are software barriers built around systems in order to make them impermeable except to people or code that is "authorized."[29] Cryptographic techniques are used to protect the integrity and privacy of information stored in computers; such techniques also protect against theft and

manipulation as information travels across networks. Some protection is offered against treacherous applets—for example, one that might reformat a user's hard drive or leak private information to the world—through security features built into Java (a computer language that greatly enhanced the range of possible Web-based engagements) that limit what applets can do. There are, however, regular announcements of flaws in this security.[30] There is fundamentally no known technical means of differentiating "good" from "bad" applets. How could there be, except in some possible future when computers would be able to discern categories of human values?

Fixing Identity

The people and institutions of the online world have diverse tastes when it comes to identification. Some are happy to link themselves to their full-blown offline identities, while others prefer to remain virtual selves. Among the second group, some are happy to maintain consistent identities represented by "handles" or pseudonyms, while others prefer full anonymity. The goal of security efforts in this category is to give more transparent access to online agents in order to stave off at least some of the threats and worries that follow from not knowing with whom one is dealing. Identifiability is considered particularly useful for recognizing malevolent or mischievous agents. And in general, it helps answer some of the questions that trust inspires us to ask: Is there a recognizable and persistent identity to the institutions and individuals behind the myriad websites one might visit? Can we count on agents online to keep their promises? For the sake of ecommerce, how do we prevent malicious agents from posing as legitimate customers or service providers and conducting bogus transactions, tricking and defrauding legitimate participants? In other words, we strive to reintroduce identifying information, at least as much as is needed to create a history, establish a reputation, hold agents accountable, and so on.[31]

Security efforts have focused on the task of making identity sufficiently transparent to protect against these and other betrayals and harms in an effort to build what the information law expert Lawrence Lessig has called "architectures of identification."[32] Mostly, they are interested in developing a strong link between a virtual agent and a physical person through a constellation of information that is commonly seen as proving identity even offline.[33] Security experts are investigating the promise for identification of biometrics—for example, fingerprints, DNA profiles, and retinal images. Furthermore, cryptographic techniques are deployed to authenticate users, computers, and sources of information by means of digital signatures and digital certificates working within a socially constructed system of certification authorities, trusted third parties who vouch for the binding of cryptographic keys to particular identities—particular persons and institutions. These same mechanisms are intended to prevent repudiation by agents of commitments or promises they may have made. A long chain of research and development focuses on various dimensions of so-called "trust management" such as these (Blaze, Feigenbaum, and Lacy 1996).

Schemes of identification, even the attenuated forms, work hand in hand with access control, because controlling access almost always means distinguishing the sanctioned, legitimate users from the illegitimate ones, not preventing everyone from using a system or the information in a system. In the case of applets, because direct examination of the applet can provide only imperfect evidence, we may rely on what is known about who sent them for another source of discrimination between "good" and "bad" applets.[34] "Trust management systems" are offered as integrated mechanisms for identifying and authenticating the identity of those people, information, and code that affect us, and they are also supposed to authenticate an applet's origins. The Snow White fairy tale offers an irresistible comparison: if Snow White had known the true identity of the bearer of the apple, she could have avoided the fateful bite.

Security experts seem to be engaged in a Sisyphean battle as they ward off attacks, repair system flaws, close up loopholes and "backdoors," and devise new layers of protection—a process that is suspended only until the next attack occurs. Outspoken security experts accept that this is an inevitable consequence of the "open" architecture of the Internet and Web, which many consider to be fundamentally insecure.[35] As a result, we live with an unstable equilibrium of relative comfort until the latest, more devastating intrusion is made public; there is a flurry of reaction, followed by relative comfort, and so the cycle continues.

Surveillance

A third layer overlaid upon the security offered through access control and transparency of identity is surveillance: we keep an eye on things in order both to prevent harms and to apprehend perpetrators after harm has been done. Surveillance can involve active watching and tracking, which can be fairly fine-grained, as demonstrated by the monitoring software that many business organizations have installed on their computer systems. Or it can be relatively coarse-grained, as are some "intrusion detection" systems, where real-time monitoring issues an alarm in response to suspicious or unusual activity, to be further investigated if necessary.[36] Surveillance can also involve passive recording (reifying) of digital trails. Popular means include logging and auditing, which creates records of activity which authorities can sift through at a later time. Logging and auditing helped authorities identify David Smith as the creator of the Melissa virus.[37]

Can Trust Be Secured?

The question is whether the array of security mechanisms—firewalls, biometrics, digital signatures, intrusion detection, auditing, and so forth—will bring about trust online. It is useful to give this question a somewhat harder edge: If we could reach an ideal end state with all three categories of security mechanisms—perfectly valid and reliable access control, identifiability, and surveillance—converging on perfection, will we have secured trust?

There is prima facie reason to answer yes because the mechanisms in question appear to address the missing triggers of trust and, failing that, to provide direct protection against some of the harms users might fear, which caused them to be distrustful in the first place. Transparent identity, for example, makes it easier to judge whether others are trustworthy—"safe bets"—or disreputable and worthy of suspicion. Mechanisms of nonrepudiation would restore accountability and restrain those inclined to dishonesty. Strong and smart walls, limits on the flow of information, and constraints on the actions that are possible, would establish safe zones by allowing in only authorized (vetted) individuals and institutions and allowing only nonhazardous actions.

Nevertheless, I will argue, in spite of its prima facie attractiveness, security—or rather the particular vision of security occupying the mainstream—will not bring about trust but, rather, surety. I argue this not because I think security is unimportant but because the ends of trust online are not well served by this mainstream vision of security. The rhetoric is misguided because when the proponents of security and ecommerce try to bind trust too closely to security, they threaten to usurp a concept as rich and complex, as intensely social, cultural, and moral as trust for one slim part of it. The mistake is not merely semantic; it has a weighty practical edge. Pursuing trust online by pursuing the complete fulfillment of the three goals of security would no more achieve trust and trustworthiness online—in the full-blown sense of these qualities—than prison bars, surveillance cameras, airport X-ray conveyor belts, body frisks, and padlocks could achieve it offline. This is the case because the very ends envisioned by the proponents of security and ecommerce are contrary to core meanings and mechanisms of trust. Security misses the mark in two ways: it undershoots trust, and it overshoots it.

Security Is No Panacea

Let us begin with the first critique of the idea of creating trust online through security—namely, that even a perfect embodiment of the three principles of security does not go far enough for trust, in significant and

systematic ways. To clarify, it will be useful to set in place a simplification, framing what is at stake in terms of "insiders" and "outsiders." Experts in computer security are worried about outsiders: malicious, avaricious, incompetent, or simply unauthorized outsiders who may break into our online space, damage or steal information, and destroy or compromise our systems. Security mechanisms are developed to keep outsiders where they belong—outside—and to help spot or identify outsiders who might be attempting to break in, in order to prevent or punish them.

This approach pays far less systematic attention to the threat of insiders, those agents—individuals and organizations—who by degrees have gained sanctioned access to our space. Some may even count among the respectable, socially sanctioned, reputable members of online society, yet they engage in actions that many citizens of the online world dislike, resent, or even consider harmful. They track our Web activities, they collect and use personal information without our permission, they plant "cookies" on our hard drives, they hijack our browsers while they download ads, they fill our mailboxes with spam, and they engage in relentless commercialism. Some of these insiders—perhaps not the "respectable" ones—"troll" our discussion groups, afflict us with hateful, inflammatory, mean-spirited emails ("flame" us), send us threatening chain mail, and even attack our virtual selves.[38] In other words, even if the walls of security keep outsiders outside, they do not curtail the agents and activities that, behind the veil of respectability and legal sanction, make online citizens skittish, cautious, and resentful. Such security barriers do not address various forms of activity that are fully capable of engendering a climate of suspicion and distrust online even if we are successful in our projects to secure the online world from "outsiders."

Even in the physical world, attention only to the threats of outsiders leaves us vulnerable to a host of dangers. In the familiar case of physical safety, some of us go to great lengths trying to protect ourselves from bodily harm, staying clear of dangerous parts of town, affixing padlocks to our doors, installing burglar alarms in our homes, accepting the ever-increasing use of video surveillance in public spaces. Homicide statistics, however, tell a curious story: when the relationship of the killer to victim is known, we find that only 22 percent of killers are strangers—the proverbial outsiders.[39] Seventy-eight percent are spouses, friends, and acquaintances. Betrayal comes from those who are allowed within our spheres of safety, within our safe zones.

My intention is not to launch into paranoid realms of suspicion and universal distrust. It is to illustrate that keeping outsiders out does not necessarily ensure safety. A wall of defense against malicious outsiders does not defend against the threats posed by sanctioned insiders, who energetically defend their "right" to exercise online freedoms—by means

of cookies, misleading registrations, matching, mining, and so on. They are, arguably, chipping away at trust just as surely as amoral hackers are. They are just as capable as hackers of causing a dangerous ebb in the abundant social capital we currently enjoy in life online.

Because it is in the nature of trust to be conservative—slow both to ebb and to grow—the results of these transgressions may not be immediately evident.[40] That the transgressions I speak of are capable of undermining trust, however, is implied by several of the works that have shaped this essay. One example is found in a long-term study of ecommerce, which shows that consumers' trust is related to their understanding of how information about them is treated; it wanes if they think that it will not be held in confidence.[41]

Another important insight that explains why interventions like the familiar suite of security mechanisms cannot fully induce trust is that trust is as sensitive to motives and intentions as it is to actions and outcomes, if not more so. It is in the goodwill of others, the philosopher Lawrence Becker has argued, that we trust or fail to trust, not necessarily in their actions.[42] As long as we believe that others are well intentioned toward us, our trusting attitude toward them will survive a great deal of bad news: "incompetence, mendacity, greed, and so forth" (Becker 1996, 51). This holds for the relation of citizens to government as well as among persons. According to Becker, only when citizens begin to attribute the poor performance of governments to deviant motivations—e.g., corruption or inappropriate power seeking—will they "respond in ways that are . . . volatile and disruptive" (Becker 1996, 59). Citizens' trust, it seems, is able to survive incompetence, at least for a while. In a similar vein, Paul Slovic (1993), an expert on risk assessment, reports that the extent to which citizens are willing to accept societal risk resulting from technological innovation is related to their degree of confidence in the motives of those in charge.[43]

Similar ideas emerge in Tom Tyler's research on public trust of police and the courts. Tyler is interested in variables that affect citizens' confidence in legal authorities, their readiness to accept outcomes, and their evaluation of the quality of decision making and fairness of procedures.[44] He finds that the most important variable for trust is the motives of authorities,[45] which he calls motive-based trust: "Motive-based trust is distinct from judgments about whether or not authorities behave as anticipated. It involves an inference about the 'spirit' or 'motive' that will shape behavior, not what specific behavior will occur" (Tyler 2001). One of Tyler's somewhat surprising findings is that in brushes with law enforcement and legal authorities, people's positive reactions are tied more strongly to inferred motives than even to whether or not the outcomes of their cases were favorable to them.[46]

The significance of these ideas to the purposes of this section is to emphasize that the behavior of many sanctioned, established, powerful

individuals and organizations is capable of undermining trust when their motives are unclear, even when the actions they undertake, such as Web tracking, for example, are not immediately aggressive or harmful. In these cases, when we learn of such activities we may find them ambiguous. What would matter to us for purposes of trust would be the motivations behind the behaviors. As long as we are not able to read people's minds, it is difficult, often impossible, to assess motives and intentions directly. So we usually find ourselves drawing on as many indirect sources as possible, sometimes resorting to subtle detection and artfulness.

One important indirect source of others' intentions is their interests. When, for example, a politician seeking office expresses concern for a particular situation, voters might attribute the expression not to genuine feeling but to an interest in being elected. In a case of this type, as much as we welcome and praise the action, it may not serve as grounds for trust as long as we see it emanating from a motive of vote seeking. In the case of Web tracking—and, more generally, information gathering and commercialism—we might initially be willing to read positive meaning into such practices. As time goes by, and we take measure of the distance between our own interests and those of the trackers (profit and potency), we begin to reinterpret those same actions as forms of betrayal. Actions that at first seem neutral or even friendly can come to be seen as sinister when interpreted in light of reasonably inferred plausible negative motives and intentions.

We all need to interact, even cooperate, with others whose interests are not consistent with our own and may even conflict with ours. In such cases, we transact cautiously, ever on the lookout for betrayal, sometimes seeking protections from the most egregious harms, betrayals, and exploitation. So trust remains elusive.

If we choose not to pursue policies for the online world that aim to contain the pursuit of avaricious interests that are contrary to those of the citizens of the Net, we are, I fear, planting the seeds of general distrust. People may continue to participate in this arena, but will do so with caution and a sense of wariness—wisely so, in interactions with those whose interests run contrary to our own, and whose actions may be annoying, bothersome, intrusive, or even threatening. Guardedness will be the norm.

Those who would pursue *security* in the name of trust do us this disservice. They focus on the outsider, the aberrant individual or organization, the trickster, the evil hacker, and the scam artist. These are the villains from whom security would protect us. In proportion to actual harm done to individuals online, too much attention is paid to the aberrant individual, the trickster, and the evil hackers lurking outside the borders of civilized online society. The media play up dramatic cases: the Melissa virus, spies who infiltrate systems and sell secrets to our enemies, or hackers who distribute unauthorized copies of intellectual

works. But these techniques to guard against malicious outsiders do nothing against agents acting behind the veil of respectability who invade our privacy and offend us by turning cyberspace to their own interests, which are not ours.

We should take greater heed of the sanctioned harms of respectable insiders; we should question the systemic imbalances between the individual citizens of the online world and the organizations that create it with little sense of the interests of the individuals. For the vast majority of Net users, it is the second group and not the first that is the significant danger; it is the second, at least as much as the first, that affects our attitudes of trust online. Powerful security mechanisms may keep us safe from malicious outsiders at the cost of our online experience, but such mechanisms still leave us vulnerable to those inside agents. We can keep out the aberrant individuals, but we remain powerless against parties that are poised systematically to exploit their positions. If we care about developing a climate of trust online—full-blown trust, not a thin substitute—we must address these conditions of imbalance between individuals and institutions. Evil hackers are not the only, nor are they the most important, barriers to trust online. If we do not address the systemic problems, trust will erode and we will not easily recover from a sense of wholesale exploitation.

Securing Trust Versus Nourishing Trust

If the earlier criticism was that security does not go far enough, this one is that security, as envisioned in the three categories, overshoots the mark and might quash trust by creating an environment in which trust is not allowed to take root and flourish. Here, an excursion back to theoretical and empirical studies of trust is useful. Trust, we learn, is an attitude. It is almost always a relational attitude involving at least a trustor and a trustee. In this relation of trust, those who trust accept their vulnerability to those in whom they place trust. They realize that those they trust may exercise their power to harm, disappoint, or betray; yet at the same time they regard those others "as if" they mean well, or at least mean no harm. Trust, then, is a form of confidence in another, confidence that the other, despite a capacity to do harm, will do the right thing in relation to the trustor. For the philosopher Annette Baier, trust is "accepted vulnerability to another's possible but not expected ill will (or lack of good will) toward one" (Baier 1986, 235); trust is the "reliance on others' competence and willingness to look after, rather than harm, things one cares about which are entrusted to their care" (Baier 1986, 259). For Russell Hardin, "[T]rust involves giving discretion to another to affect one's interests" (Hardin 1993, 507). In a similar vein, Adam Seligman holds

trust to be "some sort of belief in the goodwill of the other, given the opaqueness of other's intentions and calculations" (Seligman 1997, 43). Francis Fukuyama adds a social dimension to his account, describing trust as the "expectation that arises within a community of regular, honest, and cooperative behavior, based on commonly shared norms, on the part of other members of that community" (Fukuyama 1995, 26).

Usually trust involves more than the trustor and trustee; there is almost always an object with respect to which the trustor trusts the trustee.[47] For Annette Baier, this is demonstrated in her example of trusting the plumber to take care of the pipes in her home but not to take care of her daughter—and trusting a baby-sitter to take care of her daughter but not to take care of the pipes.[48] A person might entrust even her life to a friend, but not her heart. In the online world, there is similar discretion about not only whom one is prepared to trust but what one is prepared to entrust to them; for example, many consumers have learned that they can trust Amazon.com to deliver their orders but not trust it with their personal information.[49]

The theories of trust that I have studied differ from one another in many ways; cutting across these differences, however, is a common theme linking trust with vulnerability. When people trust, they expose themselves to risk. Although trust may be based on something—past experience, the nature of one's relationships, and so on—it involves no guarantees. As Hardin writes, trust is "inherently subject to the risk that the other will abuse the power of discretion" (Hardin 1993, 507). In trusting, we are acknowledging the other as a free agent, and this is part of the exhilaration of both trusting and being trusted. Where people are guaranteed safety, where they are protected from harm via assurances—if the other person acted under coercion, for example—trust is redundant; it is unnecessary. What we have is certainty, security, and safety—not trust. The evidence, the signs, the cues and clues that ground the formation, that give evidence of the reasonableness of, trust must always fall short of certainty; trust is an attitude without guarantees, without a complete warranty.[50] When we constrain variables in ways that make things certain—that is, safe—we are usurping trust's function. Trust is squeezed out of the picture.

No loss, some, like Richard Posner, would say: "But trust, rather than being something valued for itself and therefore missed where full information makes it unnecessary, is, I should think, merely an imperfect substitute for information" (Posner 1978). According to Posner's position, if we must choose between trust—and, consequently, vulnerability—on the one hand and certainty on the other, then certainty must win.

In practice, however, such a choice has significant consequences, which are as evident online as off. In a world that is complex and rich, the price of safety and certainty is limitation. Online as off, we do not have the

means at our disposal for assuring safety and certainty without paying this price: streamlining and constraining the scope and nature of inter actions, relationships, and community; limiting the range and nature of allowable activity; needing to make a priori judgments about those with whom one will or will not interact; having to accept increasing levels of monitoring and surveillance.[51] In general, the cost of surety—certainty and security—is freedom and wide-ranging opportunity.

The link between trust and vulnerability seems to be both conceptual and empirical. The conceptual claim is that whatever the feeling or attitude one experiences when acting and anticipating in a context of certainty and safety, it cannot be trust; this is not what trust means. The empirical conjecture, which has occupied the work of several scholars, is that in a context of complete certainty, the material conditions needed to induce and nourish trust are absent.[52] Trust does not flourish in a perfectly secure environment for reasons that are very different from the reasons for which trust does not flourish in a hostile, threatening environment. For trust to develop between an individual and either another individual or an organization, the trustor must somehow have had the opportunity to test the other agent and have had that agent pass the test. Luhmann explains the crucial role of uncertainty in the process of building trust:

> First of all there has to be some cause for displaying trust. There has to be defined some situation in which the person trusting is dependent on his partner; otherwise the problem does not arise. His behaviour must then commit him to this situation and make him run the risk of his trust being betrayed. In other words he must invest in what we called earlier a "risky investment." One fundamental condition is that it must be possible for the partner to abuse the trust. (Luhmann 1979/1988, 42)

When we are placed in a context in which we depend on others for our well-being and are assured, guaranteed by whatever means, that these others are prevented and restrained and therefore incapable of harming us, then the context, though safe and secure, is not one that nourishes trust. No test has been given; none has been passed. The variables that theorists and empirical scientists have identified as trust-inducing may signal the reasonableness of trust in a particular setting, but when grounds are transformed into guarantees of good behavior, trust disappears, replaced not by distrust but perhaps by certainty. In the presence of a violent psychopath whose limbs are shackled, one feels not trust but, at best, safety.

Another empirical basis for doubting the efficacy of security to deliver trust is that boxing people in is a notoriously bad strategy for inducing trustworthiness or even trust-reliance. Constraining freedom directly or indirectly through, say, surveillance may backfire and have

the opposite effect. Roderick Kramer, in reviewing empirical work in the social sciences, notes:

> Ironically, there is increasing evidence that such systems [based on sanctions] can actually undermine trust and may even elicit the very behaviors they are intended to suppress or eliminate. In a recent discussion of this evidence, Cialdini identified several reasons why monitoring and surveillance can diminish trust within an organization. First, there is evidence that when people think their behavior is under the control of extrinsic motivators, intrinsic motivation may be reduced. Thus, surveillance may undermine individuals' motivation to engage in the very behaviors such monitoring is intended to induce or ensure. (Kramer 1999, 591)

Philip Pettit's observations reinforce this view: "[C]ertain intrusive forms of regulation can be counter-productive and can reduce the level of performance in the very area they are supposed to affect. . . . If heavy regulation is capable of eradicating overtures of trust, and of driving out opportunities for trusting relationships, then it is capable of doing great harm" (Pettit 1995, 225).

Inducements available to individuals and institutions to encourage trustworthiness are most effective when they operate indirectly. Above all, people need to perceive that they have a choice. By means of these inducements, including sanctions and rewards, clearly articulated norms, education, character development, and so on, we may increase the incidence of trust as well as trust-reliance. If, however, we go too far and deny the possibility of choice, we deny what is fundamental to trusting relationships and climates of trust. Symbols of trust can be exhibited in small but clear ways, as illustrated at the service counter of a popular downtown cafe. A discreet sign says, "At our busy times, please be respectful of those waiting for tables." We do not coerce the good etiquette of standing in a queue, we merely make known our norms and our trust in others to behave decently.

These considerations lead me to posit two conceptions of trust. In both, trust involves placing one's fate in another's hands expecting but not being certain of the other's good will. In one conception, however, trust is understood as merely instrumental—in Richard Posner's words, "an imperfect substitute for information." Trust still has enormous value because in most cases in which we must act and make decisions, online and off, we do not have assurances; trust acts as a bridge between uncertainty and action. Yet because the ideal of these circumstances is one of assurance and certainty, we will continually work to close the gap of uncertainty, to limit and possibly eradicate the vulnerability. When buying goods online, banking, downloading information and entertainment, and so on, we welcome movement toward greater surety. Trust is the

bitter pill we must swallow in order to achieve the primary purposes of these transactions.

The other conception of trust regards it as being an overriding value, valuable in itself. When trust is so conceived, trust-based relationships would not be better if the scope of trust were minimized; rather the opposite holds. In traditional realms, trust among family members exemplifies this conception. Consider the case of parents trusting their daughter. Although they could achieve greater surety by placing their daughter under surveillance, reading her diary, checking up on her at school, and so on, most of us would agree that a trust-based relationship is better, is closer to the ideal of a parent-child relationship. The other version might provide certainty, but at a great cost to quality. In the online environment, a similar conception operates but in spheres somewhat disconnected from those we have mainly been discussing. Although no easy definition captures these spheres, they can be characterized by means of a set of typical features: they tend to reside in the not-for-profit sectors, are peer- and community-based, and are self-organizing rather than commercial, hierarchical or authority-based. They extend across a wide variety of substantive interests from political to recreational to technical, including the myriad online communities that consolidate around grassroots political interests (for example, e-thepeople and Institute for Applied Autonomy),[53] and cultural, gender, environmental, and even technology-related issues.

Reviewing specific cases, Lee Sproull and Sara Kiesler draw attention to more than six hundred volunteer health-support groups (which have been used by more than 6.5 million Americans); fifty thousand net-based volunteer technical-support user groups; software development and discussion groups (including open-source and Slashdot); mentoring and tutoring groups; political concern and advocacy communities; geographically bounded communities; peer-to-peer file-sharing communities; groups cohering around recreational online games—MUDs (Multi-User Dungeons), MOOs (Multi-User Object-Oriented Domains)—and more (Sproull 2003; Sproull and Kiesler 2003). Yochai Benkler has identified numerous highly productive efforts in what he has called commons-based peer-production, in which individuals devote time and effort to produce impressive intellectual goods online, not because a boss or manager so commands them, nor because they wish to sell the product in the marketplace, but because they choose to invest in a community-based effort (see Benkler 2002). The activities in which all these groups engage is highly diverse—the planning of civic action, promotion of political and humanitarian causes, discussion, organization, game playing, information production, protest, voting, and more. Of course, not all are directed to generally beneficial ends, including groups espousing racist, neo-Nazi, and violent anti-abortion themes.

These cases are relevant to trust because they define a realm in which we are likely to find inherently, rather than instrumentally, trust-based interactions and relationships. Unlike the relationships discussed earlier, where individuals settle for trust when they would rather have surety and safety, these relationships are valued because they are trust-based. To explore this idea a little further, consider the case of the art site TeleGarden.[54] TeleGarden is a collaborative art project and community cohering around a website and a "real" garden—roughly six square feet—which was on exhibition in Linz, Austria, at the Ars Electronica Center until November 2002. The garden is cultivated by a robot (with occasional help from museum staff) which carries out commands of TeleGarden website participants. The robot plants seeds in specified locations, waters, and provides visual feedback to the site. Created in 1995 by a team of computer scientists and engineers on the faculty of the University of Southern California, TeleGarden is now self-governed by its participants, who frequently "visit" the garden, perform maintenance, and mingle with friends on the website's discussion boards. Every few months, when the garden overfills with plants, members of the museum staff clear and replace it with fresh soil, and a new growing season is declared.

The TeleGarden project holds many dimensions of interest—its technical features, aesthetic qualities, nature of the community, why participants derive joy and satisfaction, and so on. Relevant to our discussion, however, is a particular aspect of the site's governance. Technical design features as well as explicit norms impose few constraints on participants' behaviors. Besides the rule that participants must serve conscientiously for a while, helping water the garden and participating in chat-room discussions, before they are allowed full privileges like being allowed to plant seeds, there are few restrictions. This leaves the garden vulnerable to sabotage and the community chat rooms vulnerable to offensive postings. As a matter of fact, there have been rare occasions when members have intentionally ruined the garden by overwatering and planting seeds on top of those planted by others, and have posted pornography to chat rooms. On those occasions, the organizers discussed the possibility of altering the underlying mechanisms to impose technical constraints on these forms of sabotage. They decided not to.

According to the two-conception schema, this decision indicates a commitment to trust as a defining value of the TeleGarden community. In a different version of TeleGarden, organizers might have decided to alter system software to impose watering, planting, and censorship guidelines forcibly. This would have protected the garden plot against those harms, but the community, too, would have been altered. The point is not to judge one of these alternatives to be better or worse than the other, it is to suggest that they are different. Some might prefer the

safer, more predictable version, others might find the trust-based TeleGarden more interesting, challenging, compelling, or exhilarating. In the trust-based version, part of the value and pleasure of membership is enjoying the opportunity to interact with others who behave well even though they have the freedom to act otherwise.

Although TeleGarden is only a small and somewhat obscure instance of online community, it represents a particular mode of action and relationship not uncommon online as well as off. In traditional, more widely experienced arenas, we find in ideals of human relationships, such as friendship, citizenship, community, camaraderie, family, and more, that trust is embodied not instrumentally but as an essential element. In this same way, I would argue, replacing trust with surety in many of the community, peer-based, political activities online would be to extract from them something that lies at their very hearts. How, then, does this lead us to answer the motivating questions of this article: Does surety enhance or supplant trust online? If it does, are we better off for it?

Conclusion: Struggle for the Soul of Cyberspace

I am not opposed to computer security. The basic mechanisms underlying it are diverse and capable of being shaped in an enormous variety of ways, and in turn shape the online world and the experiences possible within it. Security technology does not, in general, necessarily result in the trajectory we have just examined, in other words, does not necessarily lead to limitations on the complexity, richness, intensity, and variety of experience to be had online while not assuring protection from sanctioned predatory activities. Developed wisely, security technologies could produce a measure of safety with sufficient degrees of freedom to nourish trust. Cryptography is a good example: it can be used in the service of transparent identification, but may also be used to protect individual interests in privacy, freedom of association, and free speech.

Yet even the security mechanisms discussed and challenged here, namely, those that enable surveillance, sustain identifiability, and form selectively permeable fortresses—let us call this "high security"—are not in themselves objectionable. High security, or surety, can be good; it is even necessary for a great many settings: airplane flights, military compounds, national secrets, nuclear power plants, banks, prisons, and more are all settings where we welcome Richard Posner's vaunted certainty.[55] Nevertheless, if the arguments of this article have succeeded, they will have convinced readers that the pursuit of trust must be decoupled from the pursuit of high security; trust will not ride in on the coattails of security. Even so, these arguments do not in themselves provide an answer to the further question of what course of action we—the virtual

agents, people, institutions—who populate the online world or they, influential parties involved in building and governing the technical infrastructures, ought to pursue. That question is about what we envision for the online world, what kind of medium it is, what kind of environment it should be good for, and what values it ought to embody, trust or surety.

The social theorist Niklas Luhmann, whose profound work on trust has been widely influential, characterized trust as a mechanism for reducing complexity, enabling people to cope with the high levels of uncertainty and complexity of contemporary life (Luhmann 1979/1988). Trust makes uncertainty and complexity tolerable, enabling us to focus on few alternatives without being frozen in the mire of all possible alternatives, unable to act and decide in situations that call for action and decisiveness. In trusting, Luhmann writes, "[O]ne engages in an action as though there were only certain possibilities in the future" (1979/1988, 20).[56] Trust enables "co-operative action and individual but coordinated action: trust, by the reduction of complexity, discloses possibilities for action which would have remained improbable and unattractive without trust—which would not, in other words, have been pursued" (25).

A highly secured cyberspace provides a good climate for activities like commerce and banking, and for established commercial, public, and governmental institutions. The interests and modes of interactions that would *not* flourish in a highly secured cyberspace are likely to include the creative, political, unusual, freewheeling, subversive, possibly profane, possibly risky modes and activities of individuals. For airplane flights, we may welcome security checks, but for these other kinds of activities and interactions, for the virtual hustle and bustle that has come to resemble (and in some cases replace) much of our common experience, people avoid brightly lit scrutiny. To express the trade-off in Luhmann's terms, we may say that while both trust and security are mechanisms for reducing complexity and making life more manageable, trust enables people to act in a richly complex world, whereas security reduces the richness and complexity. Which one of these alternatives we would like to take hold online should be a matter for full and deliberate consideration and should not follow merely as an accidental consequence of immediate technological imperatives and hasty policy choices.

My own preference would be for a progressive social vision of cyberspace that preserves the degrees of freedom that trust needs. At the same time, we ought to develop technologies of security that might make possible pockets of high security for the kinds of transactions that call for it, without making that the dominant norm throughout. Outside of these pockets, we could maintain minimal protections—perhaps safety nets to prevent catastrophic harms. If we set these to be our goals, then we will have set the stage for trust. But we will *only* have set the stage. The

work of nourishing trust and trustworthiness remains as deep and complex a social challenge as ever, calling for a familiar range of diverse responses, including the promulgation of norms, moral and character education, and, ultimately, comfort for the hurt.

Notes

1. This article is an outgrowth of a collaborative project with Edward Felten and Batya Friedman, and owes much to them. It was supported by grants from the National Science Foundation, SBR-9729447 and SBR-9806234. I am grateful to my colleagues Tamar Frankel, Jeroen van den Hoven, Rob Kling, Harry Frankfurt, and Mark Poster for their probing questions and suggestions; and to Beth Kolko, Helen Moffett, Michael Cohen, and Hyeseung Song, Sayumi Takahashi, Robert Young, and Erich Deitrich for editorial and research assistance. Earlier versions of the paper were presented at the New York University School of Law Conference on a Free Information Ecology, Computer Ethics: Philosophical Enquiry 2000, in New York City in April 2000, and the Boston University School of Law Conference on Trust Relationships in Boston from September 22–23, 2003. It builds on an earlier paper, Nissenbaum (2001).
2. For a sense of the robustness of the field of research that interest in trust online has spawned, see, for example, the many and diverse projects and directions represented at the First International Conference on Trust Management (2003).
3. A misuse of language persists within the technical computer security community: proponents of a particular security device invariably use the term "trusted" to signal their faith that the system in question is trustworthy. This usage is misleading, as it suggests a general acceptance of the device in question when in fact it is the duty of the proponents to argue or prove that it is indeed worthy of this acceptance.
4. See, for example, Schneider (1999, 1): "The widespread interconnection of networked information systems allows outages and disruptions to spread from one system to others; it enables attacks to be waged anonymously and from a safe distance."
5. See, for example, Backhouse (1998, 28), discussing security issues in ecommerce; Hoffman, Novak, and Peralta (1999, 80), addressing the trust issues between consumers and businesses in ecommerce; Moskowitz (1998, paragraph 1), discussing the doubts that plague ecommerce; Ratnasingham (1999, paragraph 1), arguing that trust is an "important antecedent" for successful business relationships; Salnoske (1998, 24), commenting that both businesses and consumers regard transaction security as their biggest concern; Steinauer, Wakid, and Rasberry (1997, 118), exploring "technology or other processes that can help increase the level of confidence . . . in electronic commerce"; Woolford (1999, 18), arguing that electronic deals suffer from the problems of "authenticity and integrity"; and Camp (2000).

6. See, for example, Abdul-Rahman and Hailes (1998, 48–60), discussing the weaknesses of current security approaches for managing trust; U.S. Department of Defense (1999), classifying computer systems into four divisions of enhanced security protection; Khare and Rifkin (1997), "develop[ing] a taxonomy for how trust assertions can be specified, justified and validated"; Reiter (1996, 71), describing group communication protocols that distribute trust among a group.
7. Although I will not be discussing their work explicitly, I must acknowledge another community of researchers that has built an area of research and practice around trust, within the field of computer-human interaction. See, for example, Schneiderman (2000, 58–59), outlining certain steps, such as disclosing patterns of past performance and enforcing privacy and security policies, that designers can take to encourage trust in online relationships; Cassell and Bickmore (2000, 50–56); Corritore, Kracher, and Wiedenbeck (2001); Fogg et al. (2002); Friedman, Kahn, and Howe (2000).
8. See Schneider (1999, 12–23), evaluating whether and to what degree we can rely on existing networked information systems that support our critical infrastructures. This report urged a set of actions to increase trustworthiness and limit our vulnerability to harm, even catastrophe, that might result from failures due to malfunction or malicious attack. See also 240–55, outlining the commission's conclusions and recommendations.
9. See Seligman (1997, 19), arguing that trust in systems entails confidence in a set of institutions.
10. See Baier (1986, 259), arguing that in some instances it is more prudent to distrust rather than to trust.
11. See Becker (1996, 58), noting that a "proper sense of security is a balance of cognitive control and noncognitive stability."
12. See Seligman (1997, 16–21), explaining the difference between trust and confidence.
13. See Pettit (1995, 210), arguing that the mechanisms of trust can explain why "trust builds on trust."
14. Pettit (1995, 203), commenting that many are not proud of this trait.
15. See Pettit (1995, 203), arguing that people regard their desire for the good opinion of others as a disposition that is hard to shed.
16. See Pettit (1995, 207), arguing that "where trust of this kind materializes and survives, people will take that as a token of proof of their being well disposed toward one another, so that the success of the trust should prove to be fruitful in other regards."
17. Seligman (1997, 69), arguing that familiarity relates to the "human bond" rooted in identity.
18. See Putnam (1993, 172), arguing that reciprocity undergirds social trust, which facilitates cooperation in communities.
19. See, for example, Seligman (1997, 22), arguing that the concept of social role has been "fundamental to modern sociological analysis"; Baier (1986, 256),

arguing that people trust others to perform their roles in society; Pettit (1995, 221), arguing that divisions among people in a community are likely to reduce the chances of people from different sides trusting one another.

20. See Luhmann (1979/1988, 78–85), discussing the conditions necessary for trust to be formed; Hardin (1993, 514), asserting that the "terrible vision of a permanent underclass in American city ghettos may have its grounding in the lesson that the children of the ghetto are taught . . . that they cannot trust others"; Pettit (1995, 222), arguing that a society in which trust is found only in small family groups might become very cynical; Weinstock (1999).
21. See Luhmann (1979/1988, 84), commenting on how "complex and richly varied the social conditions for the formation of trust are."
22. See Hardin (1993, 522), discussing social mechanisms that generate trust; Pettit (1995, 220), arguing that the "trust-responsiveness mechanism" has implications for institutional design; and Weinstock (1999).
23. There is far more complexity to this issue than I need, or am able, to show here. See, for example, Nissenbaum (1999, 141), discussing anonymity and what it means to protect it; Wallace (1999, 23), offering a definition of anonymity.
24. In 1999 a thirteen-year-old boy from Haddonfield, N.J., who was participating in eBay auctions bid away $3.2 million on items like a van Gogh sketch and a 1971 Corvette convertible. His parents were successful in freeing themselves from responsibility for these transactions. See "Boy Bids $3M at Online Site," available at Associated Press Online, Haddonfield (April 30, 1999).
25. See Dibbell (1994), describing a fictional virtual rape in an online multiuser domain.
26. This is what I mean by organizing according to functionality. Structurally, a password is a very different device than a locked door, but in relation to this aspect of computer security, access control, the two are effectively the same.
27. See Ashley Dunn, "Computer World Battles Faster-Moving Viruses Technology," *Los Angeles Times*, October 4, 1999 (reflecting on the "notorious" outbreak of the Morris worm and explaining that an Internet security clearinghouse was created in response to the damage done by the worm).
28. Malware is a term used generally to refer to the varieties of computer codes produced with the intention of bothering or harming recipients or victims.
29. See, for example, Schneider (1999), defining firewalls and identifying them as one of the mechanisms used to prevent unwanted access to computer systems.
30. See, for example, McGraw and Felten (1997), reporting that a "code-signing hole" had been found in Java software; King (1996), noting that "several security flaws have been reported since Sun [Microsystems, Inc.] announced Java."
31. An interesting alternative tack is explored in work on online reputational systems that do not necessarily relay an agent's true identity. See Resnick et al. (2000).

32. See Lessig (1999, 34–35), identifying three common architectures of identity used on the Internet: passwords, "cookies," and digital certificates.
33. But see Nissenbaum (1999, 143), arguing that the capacity in the information age to aggregate and analyze the data necessary to identify an individual even without access to a name presents a new challenge to protecting anonymity, where society desires to do so.
34. The security of Microsoft's browser, Internet Explorer, is based on this principle.
35. See Schneier (1999, paragraph 9): "One problem is the permissive nature of the Internet."
36. This seems to be the form of the Federal Intrusion Detection Network (FIDNet) system proposed by the National Security Council and endorsed by the Clinton administration to protect government computers. See Marc Lacey, "Clinton Outlines Plan and Money to Tighten Computer Security" (*New York Times*, January 8, 2000), in which FIDNet is identified as part of the Clinton administration's larger computer security plan. See also "White House Fact Sheet: Cyber Security Budget Initiatives" (*U.S. Newswire*, February 15, 2000), outlining the Clinton administration's budget initiatives related to cybersecurity for fiscal year 2001; White House (2000), discussing various government intrusion detection systems. The FIDNet proposal has met with significant opposition from various civil liberties groups. See, for example, John Markoff, "The Strength of the Internet Proves to Be Its Weakness" (*New York Times*, February 10, 2000), noting that FIDNet caused alarm among civil libertarians, who said it would be used to curtail privacy on the Internet; Patrick Thibodeau (2000) reporting on privacy group's testimony before the U.S. Senate Judiciary Subcommittee on Technology, Terrorism, and Government Information.
37. See John Leyden (1999, 7), reporting that America Online assisted federal and state law enforcement agents in identifying David Smith as the creator of the Melissa virus; Lee Copeland (1999), noting that America Online tracked Smith down by tracing the virus to a list server in New Jersey; Hiawatha Bray, "N.J. Man Charged in Computer Virus Case" (*Boston Globe*, April 3, 1999), noting that America Online assisted the government agents in identifying Smith.
38. See Dibbell (1994), describing a fictional virtual rape in an online multiuser domain.
39. See U.S. Department of Justice (1994), reporting that "in murders where the relationship between the victim and the offender was known, 44% of the victims were killed by an acquaintance, 22% by a stranger, and 20% by a family member."
40. See Becker (1996, 50), noting that "ordinary life" provides substantial anecdotal evidence that most people have personal relationships in which they remain "trustful despite the known untrustworthiness of others"; cf. Slovic (1993, 677), describing trust as fragile and identifying "the asymmetry principle," whereby trust is usually created slowly but destroyed in an instant, often by a single event.

41. See Hoffman, Novak, and Peralta (1999, 82), concluding that the primary barriers to consumers' providing demographic data to websites are related to trust and noting that more than 72 percent of Web users indicated they would provide demographic data if the websites would provide information about how the collected data would be used.

42. See Becker (1996, 59), arguing that a person's loss of confidence in another person's motivations does more harm to the relationship than when the other person proves to be "merely unreliable or not credible."

43. See Slovic (1993, 680), contrasting the reactions of French and American citizens to risks associated with nuclear power and noting that the French public's acceptance of the risks is partly related to the public trust in the state-run nuclear program, which has a reputation for emphasizing public service over profits.

44. See Tyler (2001), advocating a "proactive model of social regulation" that is based upon encouraging and maintaining public trust in the "character and motives of legal authorities."

45. Tyler (2001), "Motive based trust is central to situations in which people rely upon fiduciary authorities" (366) and summary of results of an empirical study concluding that trust is an important factor in shaping people's reactions to their experience with legal authorities because "people who trust the motives of the authority with whom they are dealing are more willing to defer to that authority" and "trust leads to more positive feelings about the legal authority involved" (376).

46. Tyler (2001), "In the context of a specific personal experience with a legal authority, people are willing to voluntarily defer based upon their belief that the authorities are acting in a trustworthy manner. They infer trustworthiness from the justice of the actions of the authorities" (396), and discussion of the opportunities police officers and judges have to develop public goodwill by justifying outcomes by reference to the public's moral values, in the outcome context, and treating people fairly in the procedural context (398).

47. See Baier (1986, 236), analyzing trust as a relationship in which "A trusts B with valued thing C," and in which B is given discretionary powers with respect to C; Hardin (1993, 506): "To say 'I trust you' seems almost always to be elliptical, as though we can assume some such phrase as 'to do X' or 'in matters Y' "; Weinstock (1999).

48. Baier (1986, 245): "We take it for granted that people will perform their role-related duties and trust any individual worker to look after whatever her job requires her to. The very existence of that job, as a standard occupation, creates a climate of some trust in those with that job."

49. See Goldberg, Hill, and Shostack (2001), discussing changes to Amazon.com's privacy agreement protecting customer information that resulted in reduced protections.

50. See Luhmann (1979/1988, 20), noting that trust is based in part on familiarity, history, and past experiences, and (24), arguing that trust always involves the risk that the harm resulting from a breach of trust may be greater than the benefit to be gained by trusting; Pettit (1995, 208), arguing that irrespective of how one defines risk taking, trust always involves putting oneself in a position of vulnerability whereby it is possible for the other person to do harm to the trustor; Weinstock (1999).

51. There has been discussion in the media about the Clinton administration's proposals to monitor both governmental and private networks for signs of terrorist and criminal activity. See, for example, Robert O'Harrow, "Computer Security Proposal Is Revised: Critics Had Raised Online Privacy Fears" (*Washington Post*, September 22, 1999), reporting that civil liberties groups welcomed changes to the Clinton administration's original proposals, in particular limitations on automatic data collection; see also Tyler (2001), discussing the Clinton administration's proposal for, and reaction to, enhanced computer network security programs.

52. See Luhmann (1979/1988, 15), noting that "trust increases the 'tolerance of uncertainty' " and explaining that "mastery of events" [that is, knowledge] can replace trust.

53. See: www.e-thepeople.org and www.appliedautonomy.com/.

54. I am grateful to Gaia Bernstein for sharing the results of her research on TeleGarden, which is part of a larger book project (in progress) with Yochai Benkler, Greg Pomerantz, and Alan Toner, on commons-based productions by peers.

55. See note 56.

56. See Luhmann (1979/1988, 20), noting that trust evolves from past experiences that can guide future actions.

References

Abdul-Rahman, Alfarez, and Stephen Hailes. 1998. "A Distributed Trust Model." *New Security Paradigms Workshop* 48: 48–60.

Backhouse, James P. 1998. "Security: The Achilles Heel of Electronic Commerce." *Society* 35(4): 28.

Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96: 231–60.

Becker, Lawrence. 1996. "Trust as Noncognitive Security About Motives." *Ethics* 107(1): 43–61.

Benkler, Yochai. 2002. "Coase's Penguin, or Linux and the Nature of the Firm." *Yale Law Journal* 112(3): 369–446.

Blaze, Matt, Joan Feigenbaum, and Jack Lacy. 1996. "Decentralized Trust Management." In *Proceedings, 1996 IEEE Symposium on Security and Privacy.* Available at: http://citeseer.nj.nec.com/blaze96decentralized.html.

Camp, L. Jean. 2000. *Trust and Risk in Internet Commerce.* Cambridge, Mass.: MIT Press.

Cassell, Justine, and Timothy Bickmore. 2000. "External Manifestations of Trustworthiness in the Interface." *Communications of the ACM* 43(12): 50–56.

Copeland, Lee. 1999. "Virus Creator Fesses Up—Admits to Originating and Disseminating Melissa." *Computer Reseller News* (September 6).

Corritore, Cynthia L., Beverly Kracher, and Susan Wiedenbeck. 2001. "Trust in the Online Environment." In *Usability Evaluation and Interface Design Cognitive Engineering: Intelligent Agents and Virtual Reality*, edited by Michael J. Smith, Gavriel Salvendy, Don Harris, and Richard J. Koubek. Mahwah, N.J.: Erlbaum.

Dibbell, Julian. 1994. "A Rape in Cyberspace; or, How an Evil Clown, a Haitian Trickster Spirit, Two Wizards, and a Cast of Dozens Turned a Database into a Society." In *Flame Wars: The Discourse of Cyberculture*, edited by Mark Dery. Durham: Duke University Press.

First International Conference on Trust Management. 2003. Crete, Greece (May). Available at: www.itrust.uoc.gr/.

Fogg, B. J., Cathy Soohoo, David Danielsen, Leslie Marable, Julianne Stanford, and Ellen R. Tauber. 2002. *How Do People Evaluate a Web Site's Credibility? Results from a Large Study.* Palo Alto: Stanford Persuasive Technology Lab, Stanford University. Available at: http://www.consumerwebwatch.org/news/report3_credibilityresearch/stanfordPTL_TOC.htm.

Friedman, Batya, Peter Kahn, and Daniel Howe. 2000. "Trust Online." *Communications of the ACM* 43(12): 34–40.

Fukuyama, Francis. 1995. *Trust.* New York: Free Press Paperbacks.

Goldberg, Ian, Austin Hill, and Adam Shostack. 2001. "Trust Ethics and Privacy." *Boston University Law Review* 81(2): 407–22.

Hardin, Russell. 1993. "The Street-Level Epistemology of Trust." *Politics and Society* 21(December): 407–22.

Herz, J. C. 1995. "Cross-dressing in Cyberspace." In *Surfing on the Internet.* New York: Little, Brown.

Hoffman, Donna L., Thomas P. Novak, and Marcos Peralta. 1999. "Building Consumer Trust Online." *Communications of the ACM* 42(4): 80–85.

Khare, Rohit, and Adam Rifkin. 1997. "Weaving a Web of Trust." Available at: www.w3journal.com/7/s3.rifkin.wrap.html (accessed on November 21, 2003).

King, Richard. 1996. "Java Sun's Language Is Scoring Some Early Points with Operators and Suppliers." *Tele.Com* (May 1).

Kramer, Roderick M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50(17): 569–98.

Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace.* New York: Basic Books.

Leyden, John. 1999. "Melissa's Creator Faces 'Hard Time.' " *Network News* (April 14): 7.

Luhmann, Niklas. 1979/1988. "Trust: A Mechanism for the Reduction of Social Complexity." In *Trust and Power: Two Works by Niklas Luhmann.* Reprint, New York: John Wiley & Sons.

McGraw, Gary, and Edward Felten. 1997. "Understanding the Keys to Java Security." *Javaworld* (May 1). Available at: http://www.javaworld.com/javaworld/jw-05-1997/jw-05-security.html.

Moskowitz, Robert. 1998. "Ask Yourself: In Whom Can You Really Trust?" *Network Computing* (June 15). Available at: www.networkcomputing.com/911/911colmoskowitz.html.

Nissenbaum, Helen. 1999. "The Meaning of Anonymity in an Information Age." *Information Society* 15: 141–44.

———. 2001. "Securing Trust Online: Wisdom or Oxymoron." *Boston University Law Review* 81(3): 635–64.

Pettit, Philip. 1995. "The Cunning of Trust." *Philosophy and Public Affairs* 24(3): 202–25.

Posner, Richard. 1978. "The Right of Privacy." *Georgia Law Review* 12(3): 393–422.

Putnam, Robert D. 1993. *Making Democracy Work.* Princeton, N.J.: Princeton University Press.

Ratnasingham, Pauline. 1999. "Implicit Trust Levels in EDI Security." *Journal of Internet Security* 2(1). Available at: www.addsecure.net/jisec/1999-02.htm.

Reiter, Michael K. 1996. "Distributing Trust with the Rampart Toolkit." *Communications of the ACM* 39(4): 71–74.

Resnick, Paul, Ko Kuwabara, Richard Zeckhauser, and Eric Friedman. 2000. "Reputation Systems." *Communications of the ACM* 43(12): 45–48.

Salnoske, Karl. 1998. "Building Trust in Electronic Commerce." *Business Credit* 100(1): 24. Available at: www.nacm.org/bcmag/bcarchives/1998/articles 1998/jan/jan98art2.html.

Schneider, Fred B., ed. 1999. *Trust in Cyberspace.* Washington, D.C.: Commission on Information Systems' Trustworthiness, National Research Council.

Schneiderman, Ben. 2000. "Designing Trust into Online Experiences." *Communications of the ACM* 43(12): 57–59.

Schneier, Bruce. 1999. "Risks of E-mail Borne Viruses, Worms, and Trojan Horses." *Risks Digest* 20(2). Accessed June 17, 1999, at: http://catless.ncl.ac.uk/Risks/20.45.html.

Seligman, Adam B. 1997. *The Problem of Trust.* Princeton, N.J.: Princeton University Press.

Slovic, Paul. 1993. "Perceived Risk, Trust, and Democracy." *Risk Analysis* 13(6): 675–82.

Sproull, Lee. 2003. "Online Communities." In *The Internet Encyclopedia*, edited by Hossein Bidgoli. New York: John Wiley.

Sproull, Lee, and Sara Kiesler. 2003. "Transforming Public Volunteer Work." Paper presented at conference of the U.S. Department of Commerce, Transforming Enterprise: First International Conference on the Economic and Societal Implications of Information Technology. Washington (January 27–28, 2003).

Steinauer, Dennis D., Shukri A. Wakid, and Stanley Rasberry. 1997. "Trust and Traceability in Electronic Commerce." *Standard View* 5(3): 118–24.

Thibodeau, Patrick. 2000. "Senate Hears Objections to 'Cyberalarm.' " *Computerworld* (February 7). Available at: http://www.computerworld.com/news/2000/story/0,11280,41224,00.html.

Tyler, Tom R. 2001. "Trust and Law Abidingness: A Proactive Model of Social Regulation." *Boston University Law Review* 81(3): 361–406.

U.S. Department of Defense. 1999. "Department of Defense Trusted Computer System Evaluation Criteria." Accessed July 1, 1999, at: www.all.net/books/orange.

U.S. Department of Justice. 1994. *Bureau of Justice Statistics: Selected Findings, Violent Crime.* Volume 3. Washington: Government Printing Office.

Wallace, Kathleen. 1999. "Anonymity." *Ethics and Information Technology* 1(1): 21–31.

Weinstock, Daniel M. 1999. "Building Trust in Divided Societies." *Political Philosophy* 7(3): 287–307.

White House. 2000. "Defending America's Cyberspace: National Plan for Information Systems Protection, Version 1.0: An Invitation to Dialogue." *Executive Summary* 15.

Woolford, David. 1999. "Electronic Commerce: It's All a Matter of Trust." *Computing Canada* 25(18, May 7): 13. Available at: www.plesman.com/Archives/cc/1999/May/2518/cc251813b.html.