

# V viewpoints

DOI:10.1145/3084344

Peter G. Neumann

## Inside Risks Trustworthiness and Truthfulness Are Essential

*Their absence can introduce huge risks ...*

**T**RUSTWORTHINESS IS AN attribute that is fundamental to our technologies and to our human relations. Overly trusting something that is not trustworthy often leads to bad results. Not trusting something that really is trustworthy can also be harmful.

In many of the past 240 Inside Risks columns, we have been concerned extensively with trustworthiness, which should be a basic requirement of all computer-related systems—particularly when used in mission-critical applications, but also in personal settings such as maintaining your own quality of life. Trustworthiness is absolutely essential to the proper behavior of computers and networks, and to the well-being of entire nations and industries that rely on proper behavior of their computer-based enterprises.

### Computer-Based Systems and People

Trustworthy system behavior typically may depend on trustworthiness of

people—for example, system designers, hardware developers and programmers, operational staff, and high-level managers. Many systems that might have some assessment of trustworthiness can nevertheless be seriously compromised by malicious malware, external adversaries, and insider misuse, or otherwise disrupted by denial-of-service attacks. If such compromises arise unexpectedly, then those systems were most likely not so trustworthy as had been believed.

Thus, we need system designs and implementations that are tolerant of people who might usually be trustworthy but who make occasional errors, as well as systems that are resistant to and resilient following many other potential adversities. More importantly, we need measures of assurance—which assess how trustworthy a system might actually be in certain circumstances (albeit typically evaluated only against perceived threats). Unfortunately, some the assumptions made prior to the evaluation process may have been

wrong, or may change over time—for example, as new types of threats are detected and exploited.

In addition to trustworthiness or untrustworthiness of people relevant to their interactions with computers in the above sense, trustworthiness and specifically personal integrity are also meaningful attributes of people and governments in their daily existence. In particular, truthfulness and honesty are typically thought of as trustworthiness attributes of people. The question of whether a particular computer system is honest would generally not be considered, because such a system has no moral compass to guide it. However, truthfulness is another matter. A system might actually be considered dishonest or even untruthful if it consistently or even intermittently gives wrong answers just in certain cases—especially if it had been programmed explicitly to do exactly that. For example, such behavior has been associated with certain proprietary voting systems—see Douglas W. Jones and Bar-

bara Simons, *Broken Ballots*, University of Chicago Press, 2012.

Systems can be untrustworthy because of false assumptions by the programmers and designers. For example, sensors measure whatever they are designed to measure, which may not include the variables that should be of greatest concern. Thus, a system assessing the slipperiness of the road for a vehicle might rely upon a sensor that determines whether the road is wet. Sometimes that is done by checking whether the windshield wipers are on—which is a rather indirect measure of slipperiness and can lead to false or imprecise recommendations or actions. At least one commercial aviation accident resulted from an indirect and imprecise determination of runway slipperiness.

### **Risks of Believing in Computer Trustworthiness**

Many people believe computers are infallible and cannot lie. However, computers are created by people who are not infallible. Therefore, logically we might conclude that computers can-

not be infallible. Indeed, they cannot always perform exactly as expected, given the presence of hardware errors, power outages, malware, hacking attacks, and other adversities.

Indeed, computers can be made to lie, cheat, or steal. In such cases, of course, the faults may originate with or be amplified by people who commission systems, or design them, or program them, or even just use them, but not with the computers themselves. However, even supposedly ‘neutral’ learning algorithms and statistics can be biased and untrustworthy if they are presented with a biased or untrustworthy learning set. Unfortunately, the complexity of systems makes such behavior difficult to detect. Worse, many statistical learning algorithms (for example, deep learning) and artificial intelligence cannot specify how they actually reached their decisions, making it difficult to assess their validity.

► People who believe that online gambling is “fair” are likely to be easy victims. So can those who know it is not fair, but are nevertheless addicted (see Francis X. Clines, “Threatened

with Ruin at the Virtual Casino,” *The New York Times*, Feb. 5, 2017).

► People who believe that elections based on Internet voting and proprietary un-auditable voting machines are inherently “fair” can be easily misled. People who continue to believe that Russians had no influence on the November 2016 election in the U.S. or in the April preliminary elections in France are oblivious to real evidence in both cases. Furthermore, The Netherlands recently abandoned electronic voting systems, returning to paper ballots—wary of further ongoing Russian interference.

### **Risks of Believing in Human Truthfulness and Integrity**

Human creativity can have its downsides. For example, opportunities for ransomware, cyberfraud, cybercrime, and even spam all seem to be not only increasing, but becoming much more sophisticated.

Social engineering is still a simple and effective way to break into otherwise secure facilities or computer systems. It takes advantage of normal human decency, helpfulness, politeness,

# AD TK

and altruism. A knee-jerk attempt to rein in social engineering could involve eliminating these very desirable social attributes (which might also eliminate civility and decency from our society).

## How Does This All Fit Together?

It should be fundamental to readers of Inside Risks articles that point solutions to local problems are generally insufficient, and that we have to consider trustworthiness in the total-system context that includes hardware, software, networking, people, environmental concerns, and more. On September 22, 1988, Bob Morris (then chief scientist of the National Computer Security Center at NSA) said in a session of the National Academies' Computer Science and Telecommunications [now Technology] Board on problems relating to security, "To a first approximation, every computer in the world is connected with every other computer." That quote is even more relevant today, almost 30 years later. Similarly, all of the issues considered in this column involving computers and people may be intimately intertwined.


Science is never perfect or immutable—it is often a work in progress. Hence, scientists can rarely if ever know they have the absolute final answer. However, scientific methods have evolved over time, and scientists generally welcome challenges and disagreements that can ultimately be resolved through better theories, experimental evidence, and rational debate. Occasionally, we even find fake science and untrustworthy scientists, although these aberrations tend to be refuted eventually via peer pressure. Where science has strong credible evidence, it deserves to be respected—because in the final analysis reality should be able to trump fantasies (although this in fact may not work).

Truth is perhaps even more in flux than science, and certainly relative, not absolute—with many caveats. However, truth matters. We might paraphrase the oft-cited Albert Einstein quote as "Everything should be stated as simply as possible, but not simpler." Oversimplifications, the lack of foresight, and a seriously non-objective perspective are often sources of serious misunderstandings, and can result in major catastrophes. On the other hand, untruthfulness must not

be confused with truth, even though that confusion appears remarkably common. People who believe everything they read on Facebook, Google, Amazon, Twitter, and other Internet sites are clearly delusional.

## Conclusion

People who are less aware of technology-related risks tend to overendow computers as perfect, while computers have little respect for people. Neither computer behavior nor human behavior is always perfect, and should not be expected to be so. There are significant risks in blindly believing in computer trustworthiness and human truthfulness. We must not believe in computer infallibility, or in everything we read on the Internet in the absence of credible corroboration. But then we should also not believe people who pervasively dishonor truthfulness.

Unfortunately, the trends for the future seem relatively bleak. Computer system trustworthiness and the implications of its absence are increasingly being questioned. For example, a recent article by Bruce G. Blair (Hacking our Nuclear Weapons, *The New York Times*, Mar. 14, 2017) suggests "Loose security invites a cyberattack with possibly horrific consequences." Semi- and fully autonomous systems, the seemingly imminent Internet of Things, and artificial intelligence are providing further examples in which increasing complexity leads to obscure and unexplainable system behavior. The concept of trustworthiness seems to be becoming supplanted with people falsely placing their trust in systems and people that are simply not trustworthy—without any strong cases being made for safety, security, or indeed assurance that might otherwise be found in regulated critical industries such as aviation. However, the risks of false would-be "facts" may be the ultimate danger. An obvious consequence might be the extensive institutional loss of trust in what is neither trustworthy nor truthful. The truth and trustworthiness may be even more important now than ever before. 

**Peter G. Neumann** (neumann@csl.sri.com) moderates the ACM Risks Forum and is Senior Principal Scientist in SRI International's Computer Science Lab. He is grateful to Donald Norman for considerable useful feedback.

Copyright held by author.