Dialogues

Trust, Computing, and Society: Introduction

1

Introduction and Overview

Richard Harper

Preamble

Any glance at the contemporary intellectual landscape would make it clear that trust, society, and computing are often discussed together. And any glance would also make it clear that when this happens, the questions that are produced often seem, at first glance, straightforward. Yet, on closer examination, these questions unravel into a quagmire of concerns. What starts out as, say, a question of whether computers can be relied on to do a particular job often turns into something more than doubts about a division of labor. As Douglas Rushkoff argues in his brief and provocative book, *Program or be Programmed* (2010), when people rely on computers to do some job, it is not like Miss Daisy trusting her chauffeur to take her car to the right destination. But it is not what computers are told to do that is the issue. At issue is what computers tell *us*, the humans, as they get on with whatever task is at hand. And this in turn implies things about who and what we are because of these dialogues we have with computers. I use the word dialogues purposefully here because it is suggestive of how interaction between person and machine somehow alters the sense a person has of themselves and of the machine they are interacting with, and how this in turn alters the relationship the two have – that is, the machine and the "user." According to Rushkoff, it is not possible to know what the purpose of an interaction between a person and a machine might be; it is certainly not as simple as a question of a command and its response. In his metaphor about driving, what come into doubt are rarely questions about whether the computer has correctly heard and identified the destination the human wants – the place to which they have instructed the machine to navigate them. The interaction we have with computers lead us to doubt why a particular destination is chosen. This in turn leads to doubts about whether such choices should be in the hands of the human or the computer. The computer seems to "know" more; why

should it not decide? Thus, my use of the term "dialogue." Our interactions with computers are like those we have with people, they alter the sense we have of ourselves. Or, rather, the dialogue alters what we think we are and what we think the other is; in this case not merely a machine that acts on our command, but something greater, something in which we might come to trust.

In *From Gutenberg to Zuckerberg* (2012), John Naughton raises similarly large issues and again illustrates with destinations; but, for Naughton, we need to ask whether we can trust computing (and the Internet in particular) to lead us to dystopia or to heaven. Although the contrast he presents is not entirely without irony, heaven is represented in the duplicitous appeal of Huxley's *Brave New World* (1933), and dystopia is represented in the self-evidently bleak form of Orwell's *Nineteen Eighty Four* (1949). Meanwhile, in his *Filter Bubble* (2011), Eli Pariser complains that we cannot trust the dialogue we have with search engines. Today, in the age of "the Cloud" and massive aggregation systems, search engine providers can hide things from us in ways in which we cannot guess. When we ask search engines something, we cannot know what the answer will be, because search engine technology is now deciding what we need or want and even what is good for us to know. That this is so is at once sinister and capitalistic, Pariser argues: sinister because it disempowers humans, and capitalistic because it places the market above the public good. Search engines take you to what companies want to sell, not to what you want to know. A onetime capitalist himself, William Davidow is likewise agitated, although it is not salesmanship that worries him. We are now "overconnected," he argues in his book *Overconnected: What the Digital Economy Says about Us* (2011); we cannot trust ourselves to reason properly.

The sheer volume and scale of this discourse leads one to doubt whether any single, unified view on trust will arise from it, even if many of the authors in question want to offer one. With his highly readable *Liars and Outliers, Enabling the Trust that Society Needs to Thrive* (2012), Bruce Schneier comes to mind, as he expounds precisely such a hope. Furthermore, in addition to this list of well-known texts in the public domain, there are equally many in the more scholarly worlds of philosophy, social sciences, and, of course, computer science. In philosophy, there are an immense number of books, including Charles Ess and May Thorseth's *Trust and Virtual Worlds* (2011) or the extensive works of Luciano Floridi (e.g., 2010 and forthcoming). There are also many in sociology, including Diego Gambetta's edited collection of 1988 (which includes some philosophers, such as Bernard Williams) and, more substantively, Barbara Mitzal's *Trust in Modern Societies* (1996). Since then, there has been Piotr Sztompka's *Trust: a Sociological Theory* (2003) and Guido Möllering's *Trust: Reason, Routine, Reflexivity* (2006), and others also, too numerous to

mention. In economics especially, there has been a flowering of interest through the deployment of new experimental techniques – behavioral economics. Ernst Fehr comes to mind, with his work on the "biology of trust" (2009; see also Joseph Henrich et al.'s *In Search of Homo Economicus*, 2013). There are also a great many papers and books in computer science and human computer interaction (HCI), some of which seek to bind sociological treatments of trust to computer science terms – that is to say, to formulate sociological trust "computationally." Clark Thomborson's *Axiomatic and Behavioural Trust* paper (2010) comes to mind, as does Virgil Gligor and Jeannette Wing's *Towards a Theory of Trust* (2011). Others within the world of computer science have tried the reverse – that is, to make computational tools that are designed essentially on sociological premises, as represented in Karen Clarke et al.'s collection, *Trust in Technology* (2006). Some researchers have been even bolder, seeking to create not just summaries of computer science and sociology, but also a host of other disciplines, such as psychology and philosophy. Piotr Cofta certainly attempts this in *The Trustworthy and Trusted Web* (2011).

So what is one to make of all this? There are, I feel, five points to be discussed. First, these arguments and debates reflect and announce a historical moment, at the center of which is the concern that computer scientists have raised in the past decade or so. As Craig Mundie et al. noted in *Trustworthy Computing* (2002), the move in the late 1990s toward a computer-enabled "ecosystem" was being resisted by a public that was becoming increasingly doubtful of the trust it could invest in that ecology. Mundie and his colleagues sought to encourage a program of activities that would make the engineering of that ecosystem more robust and safe. They urged procedural and regulatory improvements that would guarantee the trust that users would need to place in computer-mediated interaction if the new ecosystem was to flourish. Before we say anything about whether those hopes have materialized, it may be worth noting that Mundie and his colleagues were not concerned with trust that users had in freestanding computers. They were interested in the connections that could be made between or through computers – in the human and business networks, in other words. As it happens, ergonomists and human factors' engineers had, in the years before, examined things such as whether the speed with which a system responded to a user command was critical to inducing or undermining trust by the user in the device itself (it is, by the way). Mundie and his colleagues were, however, interested in matters beyond the keyboard or the intricacies of "input" and "output." Their concerns resonate with the doubts that are raised by the likes of Rushkoff and Naughton. It is not the computer interfaces that matter; what matters is what computers and people do together at large. The dates of Rushkoff and Naughton's books (2010 and 2012, respectively) would suggest that the

hopes of Mundie and his colleagues have not yet been met. Nearly a decade after the publication of *Trustworthy Computing*, computer scientists are still arguing that this trust has not yet been delivered. As it happens, the solution sought by those within computer science has evolved with, for example, Gligor and Wing (2011) proposing that a formal theory that can guarantee trust in the general needs to be devised. Gligor and Wing are much more assertive about the role of computational theory in solving general doubts about trust than Mundie and his colleagues were. Be that as it may, computer scientists still think that there are big questions to do with trust, computing, and society.

It is not only computer scientists who think this; it also is not a response to their urgings that the interest of the social sciences has been piqued. Social science turned toward the topic of trust somewhat before the likes of Mundie and his colleagues. At the cusp of the twentieth century, the great French sociologist Emile Durkheim investigated the breaking down of trusted relations in books such as *On Suicide* (1897) and the *Division of Labour* (first translated in 1933); a century later, on the cusp of the twenty-first century, Barbara Mitzal argued in *Trust and Modern Societies* (1996) that the topic had come to be neglected and a renewed focus was necessary. Without such attention, the malaise about which Durkheim worried – or a form similar to it – might resurface. Concurrent with Mitzal, many economists were seeking ways of explaining apparently irrational behavior, with the question of why people trust in some situations and not in others being especially perplexing to them. This question was certainly not explicable from the utility maximization models that economists had preferred to use up to that time. The emergence of so-called institutional and behavioral economics reflected attempts to account for these ineffables. In this view, the constraints on human reasoning are sourced in such things as an aversion to rejection. This can be a determinant of choices about who and what to trust. This was the rub of Fehr's work (for example, see Fehr, 2009; see also de Quervain et al., 2004).

In philosophy, there was a similar turn toward trust and a claim that the concept had been neglected for too long. In the late 1980s, Annette Baier produced a number of influential papers arguing just that. These papers eventually appeared in her commonly cited book *Moral Prejudices* (1994) in which she suggested that trust has an affective dimension, insofar as the trusting of one person in another turns around the fact that the trusting person would, in their trust, be seeking some kind of action on the part of the trusted person. Their trust would be seeking to affect them. Onora O'Neill developed this theme in her "Reith Lectures" of 2002. Since then, a whole raft of philosophical work has emerged: the work of Ess and Thorseth (mentioned earlier) comes to mind. The same too has occurred in political science in the work of, for example,

Russell Hardin, (*Trust and Trustworthiness*, 2002) and more recently, Helen Nissenbaum (*Privacy in Context*, 2010).

One could go on. The scale of the interest in trust across the social sciences has been considerable in the past fifteen years or so, just as it has been within the world of computer science. However, no end would be served by trying to list and taxonomize all these efforts at this point in our concerns. But what is important to note, and this is the second of my five points, is that this combination of interests not only asserts the connection between computing, trust, and society, but it can result in a blurring of the differences in these concerns. When we try to navigate to issues of trust, it is often not clear what we are looking for. Are we looking for trust in some general sense or in some particular way, related specifically to computers? Or is it trust in some bigger sense, in relation to the regulatory frameworks in which we use computers? Or, for a third possibility, is the concern related to some very particular disciplinary focus? Computer scientists have a strong interest in cryptography, for example, which is not the same as the philosopher's interest in conceptual clarity, even if both concerns seek to deliver trust – trust in a system in the first case, trust in understanding in the second. It is often quite difficult, however, to identify issues when a confluence of disciplinary interests conflates topics and agendas. Just as it might be true to say that issues of trust, computing, and society are multidimensional, it is also true to say that the dimensions in question are all too easily confused when the various perspectives of different disciplines are thrown in.

This leads to the third point. If it is the case that various disciplines turned to issues of trust at a similar time, it needs to be recognized also that the differences have led to diverse treatments. Thus, although various disciplines might announce a concern with trust, one must not be tempted to think that their approaches, considerations, and insights will all fit together. They might fit together in some respects, but care is required; often they do not. Additionally, part of the value that might be found here has to do with these very differences, which might be lost through integration and merging. Differences can produce a wealth of reasoning – greater coinage for the mind, if you like. Melting that coinage does not necessarily increase the volume or the value. It can, in fact, result in less. It can debase the coinage. The situation is not helped by the rhetoric of many of the disciplines involved, a rhetoric that can resonate with rather gross claims – such as the science in question "uncovering the truth" and "getting to the reality." There can be no doubt that disciplines (often) get to views that accord with these claims, but this needs to be understood within the frame of the inquiries in question. One needs to be sensitive to the significance of perspective, methods, and topics between disciplines. One needs to

be open to the possibility that when, say, sociologists look at trust, the phenomenon of interest to them may not easily fit into the concerns of, let us say, an economist. It is similar with the earlier example about philosophy and computer science, illustrated with their respective concerns for cryptography and conceptual analysis. These concerns are of two very different kinds, yet both disciplines are seeking assurances of one kind or another. The important point is that, although both consider trust, they do not do so in equivalent ways.

This leads to my fourth point, which questions whether trust is something that can be treated with the "scientific method" as is implied in some of the debates and misunderstandings alluded to previously. Trust is, after all, an idea – a state of mind, if you like – and not a physical phenomenon. It is hardly surprising then – despite all the debates, the papers, and the calls for action – that trust has a vitality in everyday life that makes the concept robust; how the concept is used in the traffic of living remains pretty much how one would expect it to. People don't find their understanding of trust altering in light of these debates; only the applicability of the concept shifts. What was once trusted may now not be, for example. How to use the concept remains the same regardless; its use is common sense, if you like, a skill constitutive of competent language use. As Peter Winch noted long ago in his *The Idea of a Social Science* (1958), concepts like trust are best thought of as philosophic in nature. Inquiries into their nature are akin to studies into linguistic practice and use; the purpose of such inquiries is to offer perspicuity and clarity when that is required – as it is today in regard to trust, when technology and trust are paired in ways that can be confusing as well as enlightening. Here I must pay heed to my own intellectual roots, not so much in Winch as in the philosophical investigations of Ludwig Wittgenstein and Gilbert Ryle (in such works as the *Philosophical Investigations* (1953) and *The Concept of Mind* (1949)).

It is also hardly surprising that when one looks at the immense literature on trust, computing, and society, one sees a raft of inquiries that rely on the concept so as to ask focused questions. It does not matter that in some instances these questions are large and worrying, as we saw with Pariser and Naughton. These authors are not asking what trust means in any perplexingly radical ways. They do not need to suggest redefinitions of how trust is to be understood or used. All they are trying to say is that questions of trust apply more than we might think in our everyday interactions with computers. Pariser and Naughton are making judgments about scale, if you like, about how much trust there is (or isn't), or where trust is absent when one would expect it to be present. They are not seeking to alter how the concept of trust is to be used, only where and when; theirs is a concern with empirical matters, not conceptual. In these and other cases, one does not find that the concept of trust is altered. Its meaning

remains taken for granted. As Wittgenstein and Ryle showed so clearly, and, in their wake, as Peter Winch also, any attempt to explore concepts used in everyday life must depend in the first place on vernacular competence with those concepts. Therefore, what the various inquiries into the topic of trust achieve has not been, in my mind, redefinitions of trust; instead, the outputs of these inquiries relate to questions of particular interest given the meaning of trust as it is already known to be. When I read claims by various authors that they are offering definitions of trust, I am often skeptical; I am suspicious that they hide the fact that they are using the term in everyday ways even as they claim to do otherwise. One might allow this dissembling if it leads to a useful nuance and focus for some inquiry, but one ought to balk at such claims, especially if these arguments end with directives about how the concept "ought to be used." This distracts from the value that can be found in their studies, their conceptual vanities notwithstanding.

Key to judging any and all inquiries on the topic of trust, computing, and society is, it seems to me, to evaluate findings in terms of the proper use of perspectival and methodological constraint. These constraints should not confound sensible use of the concept. An analogy might be helpful here. It seems to me that, in its use, the concept of trust has similarities with the concept of truth. When one uses the concept of truth, one does not ask what the concept means (one does not ask "what is truth?"), for if one asked that, one would not be able to deploy the concept. Instead, one asks what is true in some particular context. One asks what is true "here" or "here." This allows us to look at evidence, at what is relevant, and at what is a reasonable judgment in some situation. And I think the same holds for trust. Trust is a concept that allows us to make judgments, to call attention to issues, and to account for choices of various kinds. But this use only succeeds if it starts with (or from) competent understanding of how to use the concept in the normal traffic of living.

My fifth point follows directly from the fourth. In one important respect, the concept of trust is not like the concept of truth, because truth suggests an orientation of neutrality – a calmness, if you like. But a property of the way the concept of trust gets used is to cause agitation. Leaving aside what might be addressed when the term is used – questions of scale and connection mentioned earlier, for example – the use of the term can also cause (and lead to) worry. In many cases, of course, such worry is needed and indeed rightly sought. Part of Nissenbaum's book *Privacy in Context* (mentioned previously) is utilized precisely to make people more concerned than they are. She uses the term "trust" as a method to negate complacency. In this respect, the concept can be a lightning rod, which can come at a price. It is not always sensible to use

the term "trust" in this way. Often, this use results in trust pushing other ways of thinking about an issue out of the mind; it can somehow come to dominate explanation. In my view, use of the term "trust" can sometimes undermine more sensitive and careful deliberations in which other concepts are more applicable.

Some of this edging out of other concepts is quite subtle. When one looks at the literature of trust, one becomes very aware of how little attention is given to something that would be ironically central to these concerns: that is to say, what happens when trust does occur. The literature is really about *mistrust*. The use of the term "trust" does not lead to examinations of trust, then; on the contrary, much greater attention is given to those situations when trust is *absent*. Richard Holton's (1994) response to Baier's (1994) analysis of the affective in trust comes to mind; neither Holton's nor Baier's study is really interested in normal affairs, in activities in which trust is taken for granted. Their focuses are instead on those situations in which trust is *not* present; they seek to reconstruct what normal affairs might be in that light. This seems a peculiar way of doing business, the kudos Holton's work currently receives notwithstanding.

In other words, the concept of trust can both point to – lead the eye to see – one set of issues and make other issues disappear from view. Given what I have already said about the conflation of disciplinary perspectives, this is, I think, a real concern. One needs to be wary when the term "trust" is used: wary of what we are not seeing as much as what we are being led to see.

Overview of the Book

It is in light of these considerations that the following collection has been brought together. The arguments have been selected to provide a balanced set of perspectives. Each articulates a different purpose, or proposes a different emphasis given its disciplinary starting point. The overall goal of the collection is to provide the reader with a sense of – and perhaps a sensibility for – the overall topic that these views represent: the topography of arguments about and issues related to the general question of trust, computing, and society. Part of this sense will consist of the ability to historicize the arguments in question, to see where they come from, what they are trying to do, and where they are trying to go. And part of gaining this sensibility has to do with appreciating the difference between analytical and theoretical discussions about trust and society and arguments and considerations that are primarily pragmatic, as when design and engineering are at hand. As will be seen, when principal turns into practical choices, it does not mean that principal disappears or that the theoretically deduced concerns are lost. They remain but often in different forms, often with