Information **2011**, *2*, 195-216; doi:10.3390/info2010195

OPEN ACCESS

information

ISSN 2078-2489

www.mdpi.com/journal/information

Article

Trust, Privacy, and Frame Problems in Social and Business E-Networks, Part 1

Jeff Buechner [1,2]

[1] Department of Philosophy, Rutgers University-Newark, Newark, NJ, USA;
E-Mail: buechner@rci.rutgers.edu

[2] The Saul Kripke Center, CUNY, The Graduate Center, New York, NY, USA;
E-Mail: JBuechner@gc.cuny.edu

Received: 24 December 2010; in revised form: 21 February 2011 / Accepted: 22 February 2011/ Published: 1 March 2011

Abstract: Privacy issues in social and business e-networks are daunting in complexity—private information about oneself might be routed through countless artificial agents. For each such agent, in that context, two questions about trust are raised: Where an agent must access (or store) personal information, can one trust that artificial agent with that information and, where an agent does not need to either access or store personal information, can one trust that agent not to either access or store that information? It would be an infeasible task for any human being to explicitly determine, for each artificial agent, whether it can be trusted. That is, no human being has the computational resources to make such an explicit determination. There is a well-known class of problems in the artificial intelligence literature, known as frame problems, where explicit solutions to them are computationally infeasible. Human common sense reasoning solves frame problems, though the mechanisms employed are largely unknown. I will argue that the trust relation between two agents (human or artificial) functions, in some respects, is a frame problem solution. That is, a problem is solved without the need for a computationally infeasible explicit solution. This is an aspect of the trust relation that has remained unexplored in the literature. Moreover, there is a formal, iterative structure to agent-agent trust interactions that serves to establish the trust relation non-circularly, to reinforce it, and to "bootstrap" its strength.

Keywords: trust; akrasia; Deweyan conception of knowledge

1. Introduction

Artificial agents are ubiquitous in the context of Internet activities, and, most likely, they will only increase in frequency and magnitude in the future. Private and public information is disseminated in the web, and although there are various formal and informal structures in place that serve to safeguard that information, the worry is that such information will become available to anyone. The consensus view is that cyber piracy is rampant in the world. Cyber warfare and cyber espionage are deep concerns for both world leaders and the general public. The Obama administration estimated that corporate cyber espionage resulted in one trillion dollars loss of intellectual property last year [1][i]. The irony (at least in the United States) is that the means by which cyber espionage is dealt with results in a loss of protection of privacy for individual citizens. Privacy in the context of the Internet concerns private information about an individual. Clearly, there are questions about whether certain information is genuinely private, just as there are important and vexing issues about what constitutes privacy and violations of privacy, but these issues will not be our concern here. For one, we simply assume (for sake of argument) that all cases of private information are cases where the information is genuinely private.

My project is to examine an aspect of the trust relation that has been neglected in the literature—*i.e.*, the way in which a trust relation can conserve valuable computational resources. In e-networks, two different kinds of trust relations are at work—a moral and an epistemic trust relation. They are often in considerable tension, especially with respect to the conservation role of trust relations. That is, protecting privacy on the Internet, where artificial and human agents interact, requires trust among agents. However, establishing knowledge claims also requires trust among agents (on a social conception of knowledge). In some cases, epistemically trusting other agents may preclude morally trusting those same agents, and conversely.

My overall project is divided into two parts. In the present paper—which is Part One—I will articulate several problems that arise for the notion of epistemic trust in the context of privacy issues in e-networks. In another paper—Part Two—I will elaborate on the analogy between frame problems and trust relations and describe several problems arising at the intersection between epistemic and moral trust in the context of privacy issues in e-networks. In particular, I will describe the formal, iterative structure of agent-agent trust interactions that serves to establish the trust relation non-circularly, to reinforce it, and to "bootstrap" its strength.

Where human agents and artificial agents interact in contexts in which trust and privacy are in play, there are several different levels at which those interactions can be described. Moreover, depending upon the degree to which artificial agents simulate human agents, and depending on the properties which are simulated, there are different aspects of trust and of privacy that come into play. This means that there cannot be, at the present moment, a uniform account of trust and privacy in e-networks that will spell out necessary conditions that will be applicable to any future e-networks in which trust and privacy are salient. For instance, if artificial agents are fairly primitive in terms of their mental properties, then human agents might not have a trust relation with them, but, rather, simply think of them in terms of their reliability and how well a human agent can predict their future behaviors in various contexts[ii]. On the other hand, if an artificial agent is mentally sophisticated, then it might be that a human agent can have a trust relation with it. With respect to the issue of privacy, it is important to be cognizant of these potential differences, since it might be the case—as we will discuss below—that for primitive artificial agents, an algorithm could guarantee certain forms of behavior that respected privacy of information, while for sophisticated artificial agents those algorithms might be infeasible, and it would be better for human agents to have trust relations with those sophisticated artificial agents.

An additional layer of complication arises owing to the nature of the transactions and processes intrinsic to e-networks—viz., that both moral and epistemic notions of trust are required. Epistemic trust is less well known than moral trust, but it is no less important in e-networks than the notion of moral trust. The reason is simple: No single agent—human or artificial—can determine the reliability (much less access) all of the information that flows through a sufficiently large e-network. Thus, each agent must trust that other agents either have or send reliable information. Typically the notion of epistemic trust surfaces in discussions of knowledge in contexts where many individuals are involved, but even if the status of a belief arising in agents in e-networks is not required to be knowledge, but simply belief (perhaps of such-and-such a degree), the role of epistemic trust is still active, since acquiring information for a belief is best done where the information comes from a reliable source. The problem that is created by epistemic and moral senses of trust in e-networks where privacy is a central concern is that the two notions of trust can conflict with one another. For instance, epistemic trust might require that agent A epistemically trusts agent B, while moral trust might require that agent A not trust agent B, since agent B, though epistemically reliable, is not morally trustworthy. As we shall see below, there are other ways in which the notions of moral and epistemic trust can conflict in e-networks, which are quite subtle and not easy to diagnose (or, indeed, even to recognize).

2. Frame Problems, Conservation of Valuable Computational Resources, and Trust Relations

The general form of a frame problem is that to achieve reliability in reaching some goal, an agent must do α, but to do α requires an infeasible amount of valuable computational resources—β—which the agent does not have. The problem is how the agent can achieve reliability in reaching their goal without using β resources (and so, without doing α)[iii]. There is an analogy with problems in computational complexity: How can problem X be solved without using β resources, where solving X requires—because it is the complexity profile of the problem—β resources. In cases where a computational system does not have the resources to solve a problem requiring those resources, what sorts of solutions are available to the agent, and how reliable are they? Some solutions allow a computational system to do all of the steps necessary for solving X without expending the resources (such as DNA computers), while other solutions relax somewhat the requirements on reliability. Examples of these include probabilistic solutions and approximation solutions to X.

Although there are several different kinds of frame problems, all of them have the general characteristic sketched above: Achieving reliability in reaching one's goals (of such-and-such a kind) appears to be computationally intractable. However, humans are able to achieve reliability and achieve their goals (of such-and-such a kind) in many different kinds of situation (though certainly not all) without doing what is computationally intractable. How is it done? One particular kind of frame problem is how humans are good are inferring what does and does not change in one's local environment when various actions are performed. How can a human being be reliable in doing this—

after all, wouldn't making these inferences reliably require that humans keep track of each thing that changes and each thing that does not change when an action is performed? However, if there are M objects and N properties, there are M × N possible ways in which the local environment can change and there are, similarly, M × N possible ways in which the local environment stays the same. An algorithm for determining what does and does not change when an action occurs that (i) explicitly records each object and property pair, for every object and every property, and (ii) checks reality against each pair, would consume too many computational resources (especially if it either needs to be updated every, say, millisecond or needs to take into account multiple inter-dependent changes in a given temporal interval). The problem at hand, then, is how to quickly and reliably determine what does and does not change when an action occurs without performing (i) and (ii).

An important question is whether there are informal counterparts to these concepts in the moral trust relation (as distinguished from the epistemic trust relation, where—as we shall see—those informal counterparts also occur). There is a good deal in the literature on the concept of moral trust that suggests that one function of the trust relation is to provide a way of side-stepping a computationally intractable problem, though no one has explicitly claimed that such a feature is an essential feature of the trust relation, let alone explicitly recognizing it. Annette Baier has made the claim that in a trust relation we entrust something—a thing, a task, a psychological feature—to the care of another. Where A cares about something, and A trusts B, A trusts that B will care about it as well. However, in order for B to care about something that A cares about, both A and B must exhibit a certain form of discretion. That discretion is necessary for proper caring. Indeed, Baier claims, especially in her essay "Trust and Anti-trust", [2] that it is not feasible to explicitly articulate, in advance, just exactly how the caring—on both sides of the trust relation—is to be administered. If both agents had to know the specific details of how caring proceeds in all possible circumstances before they could engage in a trust relation, the trust relation could not take place, since it would be a computationally infeasible task to acquire that knowledge (for the number of distinct possible circumstances is unbounded). Notice that it is part of the trust relation that explicit listing of the conditions for caring in all possible circumstances is not needed.

Baier's account of the trust relation in terms of caring has been criticized by Margaret Urban Walker [3] as not saying enough about those aspects of caring that cannot be explicitly articulated in advance. Walker thinks that it is also not possible to articulate in advance what one is entrusted to care about. Whether Walker is right about this, though, is not essential to my point, since even if it is only providing for the caring relation that cannot be explicitly articulated in advance, it is still the case that the trust relation allows one to have a relation where a computationally infeasible task, that might be thought to be necessary for the trust relation, cannot be performed (because it is computationally infeasible). What has not been noticed in the literature on trust is that this very feature—the trust relation renders superfluous a feature that might be thought necessary to establish a trust relation—should be taken to be an essential feature of a trust relation. (I shall not argue this here, for it would require a separate section[iv].)

Walker claims that "[t]he language of entrustment here is a misleading model; it is more accurate to say that I trust her to be honest or to be loyal, that is, to behave in ways that satisfy a general norm or standard that poses different [kinds of] demands and [that] calls for distinct judgments in different contexts. Or again, what can we say we are entrusting to all others when any of us walks down the street without concern? Here there is not one norm presupposed but indefinitely many, any of which might be identified only in the breach or under threat, and each of these norms requires many things we cannot enumerate in advance."[3] If we had to enumerate in advance all of these features in order for the trust relation to be satisfied, we would not be able to do so either optimally or even sub-optimally, since to do so would require more resources than we have. However, if it is a feature of trusting the other that we trust them—and they trust us—to do the right things when the circumstances demand that, then it is certainly an essential property of trusting that we are able to avoid having to perform tasks that require too many resources to be satisfactorily performed.

A notion of trust put forth by Trudy Govier has several features, one of which is that an agent who trusts another agent has a disposition to provide a positive interpretation of the actions of those whom that agent entrusts. [4] This is an important way of saving valuable resources, since it would be difficult for any given agent to continually monitor the actions of another agent in the service of evaluating them with respect to several different dimensions of evaluation. Govier also thinks that entrusting agents in a trust relation attribute a sense of general integrity to entrusted agents. This, too, is a means of saving valuable resources, since it would be difficult, if not impossible, for an agent to continually re-evaluate another agent as a whole with respect to integrity. The difficulty is that integrity is a capacity of an agent—in the sense of being a virtue—as much as it is a disposition to behave in a certain, morally evaluable way. Consequently, monitoring it is difficult, since it will, for example, fail in some contexts or be difficult to recognize in other contexts.

Govier also brings the notion of expectation into the picture of trust, but in an interesting way—she takes it to be a feature of trust that trusting agents have expectations that entrusted agents will exhibit good behavior, where the trusting agent's expectations are based solely on beliefs about the entrusted agent's motivation and competence. Here, too, resources are conserved, since having an expectation that an agent will behave in a certain way sidesteps the need to actually observe the agent's behavior from time to time (or at certain times reflecting a good statistical sample). The idea that an expectation plays such a role in the trust relation has never been explicitly developed, but it is clear that it is central to the notion of trust. Moreover, although positive expectations of behavior on the part of the entrusted do not singly constitute the trust relation, it is important to see that those expectations—and the role they play in conserving resources an agent possesses—are justified when the normative notions that flesh out the notion of trust are brought into play.[v]

Govier also sees a need to bring into the trust relation the idea that a trusting agent accepts that there is risk in trusting and that this risk leaves the agent vulnerable in certain ways. However, it is not the case that the trusting agent actually computes a risk profile, or provides an exhaustive list of the ways in which she is vulnerable. Rather, the trusting agent merely acknowledges that she is at risk or is vulnerable, without developing additional justifications for this contention by listing the ways in which that can occur. Indeed, it would be inimical to the trust relation if the trusting agent *did* provide risk profiles and exhaustive enumerations of vulnerabilities. The point of the trust relation is that such requirements are relaxed, if not totally abandoned. Govier's conception of trust is well-suited for the kinds of activities that occur in e-networks, where agents—human or artificial—trust other agents in the same way that a human agent will trust an elected official or will trust the professional advice that is given by doctors or lawyers. (Walker has criticized Govier's conception of trust, alleging that it does not fit all of the cases that an adequate conception of trust should fit. However, our goal here is not to develop a notion of trust that fits all of the salient cases, but rather to point out that there are certain features of trust—previously unnoticed in the literature—that are pivotal in thinking about the ways in which trust and privacy interact in e-networks.)

The idea that in the trust relation there are constraints on interpreting the actions of entrusted agents has been considerably developed by Karen Jones[5], who focuses on the ways in which unfavorable interpretations of another's actions are jettisoned and, in their place, favorable interpretations offered. Moreover, the range of favorable interpretations is kept quite small, so that the entrusting agent will not have to survey different kinds of favorable interpretations—and, presumably, ascribe to the entrusted agent various properties based on that interpretation. In this way, an agent saves valuable resources. Obviously, there are circumstances in which an entrusting agent must give up a favorable interpretation. However, these circumstances are those where the behavior of the entrusted agent is out of the ordinary, or abnormal. Just as in frame problems in artificial intelligence, there are circumstances where the typical behavior of an agent (or a process, *etc.*) fails to occur, so, too, there are circumstances in trust relations where an entrusted agent acts abnormally (with respect to the normative expectations of the entrusting agent). It is a hard question in non-monotonic reasoning (which provides a formal class of solutions to frame problems) as to how an agent is able to detect such abnormalities without having to explicitly list all normal and abnormal conditions and then monitor each situation with respect to this list for evidence of an abnormality. (Indeed, how human agents are able to reliably detect abnormalities without explicitly monitoring them is an open, "hard" problem in artificial intelligence, as well as in cognitive science.) The correlative question for a trust relation is how agents engaged in such relations are able to determine when the relation should be broken—that is, are able to determine when conditions are out of the ordinary. If agents cannot tell when conditions are out of the ordinary, it is not clear that they fully understand what it is to commit to a trust relation.

Judith Baker has argued that entrusting agents may believe that an entrusted agent is worthy of trust, even where there is evidence that this is not so.[6] This is a difficult topic that is important—since there is a relation between epistemic trust and moral trust—and this point has not received the attention it deserves. Where agents do not rely upon the best evidence in making decisions, they can be accused of acting irrationally in an epistemic sense. But can they be accused of acting irrationally in a moral sense? This is not a case of epistemic akrasia (where an agent acts irrationally in the view of an epistemic norm, even though, in the view of another (less binding) epistemic norm, the agent acts rationally; rather, it is an instance of "hybrid akrasia," where the agent acts irrationally in the eyes of an epistemic norm, but rationally in the eyes of a moral norm[vi]. The point, though, is that even where there is evidence against trusting another agent, one trusts it nonetheless. This, in turn, shows that there are regions in the trust relation where it is not clear whether an agent should trust or not trust another agent, but that the agent should, unless there are clear indications to act otherwise, trust the other agent. Thus, such aspects of the trust relation are second-order aspects—that is, the agent knows that she is in a region where trust might be given up since there is epistemic evidence that the entrusted cannot be trusted. But, reflecting on the trust relation, she affirms her trust in the entrusted agent on the grounds that it is part of the trust relation to trust the other agent even where there are certain grounds for giving it up (though the grounds cannot be conclusive, whatever that might mean in this context).

The point of the discussion in this section is to illustrate the ways in which a trust relation is resource conserving. As already mentioned, this is an aspect of the trust relation that has not been explicitly noted in the literature, even though there are, in almost every discussion of trust and what it consists in, ample descriptions of features of the trust relation that are clearly resource conserving. I will now use this feature of the trust relation to describe the ways in which trust and privacy interact in e-networks, and provide recommendations for how one should think about the trust relation in such situations.

3. Epistemic Trust, Social Conceptions of Knowing, and Conservation of Computational Resources

In his important and groundbreaking paper "The Role of Trust in Knowledge," John Hardwig [7] argued that in many situations there cannot be knowledge unless there is epistemic trust. This claim shattered the view that trust is inimical to knowledge, since only epistemic justification of true beliefs (and whatever else is needed to deal with "Gettiered" beliefs[vii]) counts as genuine knowledge. Hardwig instituted the social conception of knowledge (which is not to be confused with views that see knowledge as a social phenomenon) and this view of knowledge contrasts with an individualistic conception of knowledge. He describes various situations in which there are either (i) many different people, each with knowledge such that no one person can acquire all the knowledge from each of them and such that all of the knowledge is needed to produce one further item of knowledge or (ii) many different things necessary for, say, performing an experiment, such that no one person can perform all of them. In such situations, there cannot, at least on an individualistic conception of knowledge, be any knowledge, since there is no one person who possesses all of the ingredients that are necessary for knowledge. However, on a social conception of knowledge, there could be knowledge. Hardwig has noticed that in situations where human agents do not have the resources to acquire knowledge, the epistemic trust relation is necessary for acquiring knowledge. Notice that, for Hardwig, it is not the case that epistemic trust is used to conserve computational resources (even though this is, indeed, true), but, rather, that in situations in which an agent does not have the computational resources to achieve knowledge, trust is needed to bridge the gap. Here, too, the idea that trust conserves computational resources is implicit, though not explicitly acknowledged.

The question is how to spell out the necessary conditions for social knowledge. The necessary conditions for individualistic knowledge are simply the necessary conditions for knowledge, period. The critical point is that (in many situations), as Hardwig says, "those who do not trust cannot have the best evidence for their beliefs. In an important sense, then, trust is often epistemologically even more basic than empirical data or logical arguments: The data and the argument are available only through trust. If the metaphor of 'foundation' is still useful, the trustworthiness of members of epistemic communities is the ultimate foundation for much of our knowledge."[7] It might even be argued that data is not data, until trust is in place, since taking something to be an item of data will, in certain kinds of situations, require that one trust many others who acquire the information that becomes the data through experimentation and interpretation of those experiments. A number, by itself, in the context of an experiment, is not a datum until it has been interpreted by a member of the scientific community performing the experiment. An important problem whose solution is necessary for developing the social conception of knowledge is to categorize the kinds of situations in which the social conception operates (and for which the individualistic conception is not appropriate), though there has not been much work on this problem in the literature.

Hardwig describes two different kinds of situations—roughly adumbrated above—in which the social conception of knowledge is appropriate, although there may also be additional situations, which are qualitatively distinct from those two. In the context of information systems and e-networks, this is an important task. The reason why this is so is that epistemic trust is critical to e-networks in which there are a great many agents—both human and artificial—none of whom have access to all of the data which flows through the network, and where knowledge depends upon the union of many distinct data items, not all of which can be possessed by any given agent in the network. Although there are certainly situations in e-networks in which the individualistic conception of knowledge is appropriate, there are also situations in which the social conception is appropriate and the individualistic conception is not appropriate. Certainly, individual agents cannot possess a single piece of data throughout the history of that datum in an e-network, since (i) agents usually acquire other data throughout a given e-network history; (ii) agents do few other tasks—*i.e.*, they are specialized for certain kinds of jobs; and (iii) there would be a collision of data items in the e-network, if agents held onto single data pieces and attempted to accumulate other data items.

Certainly, a good distributed algorithm will determine which agent acquires which data (and for how long) through a computation[viii]. However, in e-networks, the historical course of any given item of data is not entirely determined by an algorithm, since there will be sufficiently many external opportunities—that is, opportunities that are recognized by the algorithm, but not explicitly controlled by it—in which data may reside with an individual agent for longer or shorter periods of time, depending upon circumstances. The point is that agents—both human and artificial—in e-networks must depend upon other agents in order to perform their tasks successfully. There is little difference between these kinds of situations and the ones that Hardwig describes, in which, say, there are so many human experimenters (he uses the example of an experiment in theoretical physics in which the lifetime of charm particles is measured) required for the success of the experiment such that no one experimenter can perform all of the tasks required for the experiment to succeed. Thus, the epistemic situation in e-networks nicely corresponds to the epistemic situations in, for instance, large experiments in particle physics, or in collaborative activity in proving theorems in mathematics that require information from several disparate parts of mathematics (such as number theory and algebraic topology). For that reason, Hardwig's work is directly relevant to epistemic trust relations in e-networks where privacy issues are paramount.

One way in which a necessary condition for knowledge on the social conception of knowledge is fleshed out is by the clause: "S_1 knows that S_2 knows *p*". However, the problem engendered by this clause is that there will be situations in which no one (other than the "knower") knows who knows enough. Another (indeed, "fatal") criticism of this clause is that S_1 will have to know *p* in order to determine that S_2 knows *p*—in which case, it is otiose for S_1 to depend upon S_2 in order to know *p*. Notice that even where S_1 does not have to know *p* in order to know that S_2 knows *p*, it might be computationally infeasible for S_1 to determine that S_2 knows *p*, if S_1 has to know who knows *p*. This is especially so in an e-network, where determining who knows *p* might require solving graph problems that are in the complexity class NP (or in higher complexity classes). How would a notion of epistemic trust serve to eliminate performing a computationally intractable task on the part of any given agent in an arbitrary e-network? If one agent trusts that another agent knows *p*, then there is no need to, for instance, determine who "knows *p*" in the e-network. (Of course, there must be some reason to think that there is someone who knows *p* in the e-network.) So there are two ways in which there are computational resource savings in e-networks: (i) not having to determine who knows *p*, and (ii) not having to justify that *p* is true.

What are the necessary conditions, in the social conception of knowledge, for knowing *p*, given that "S_1 knows that S_2 knows *p*" will not work? Hardwig's proposal is that the necessary condition for knowing *p* should be: 'If S_1 knows that S_2 knows *p*, then S_1 knows *p*.' What is the difference between the rejected clause and the emended proposal? It is that S_1 knows *p* when the condition that S_1 knows that S_2 knows *p* has been satisfied. But what about the criticism that the original clause is too strong—that it requires that S_1 knows *p* in order to know that S_2 knows *p*? Here it is necessary to examine the conditions under which S_1 could come to know that S_2 knows *p*, without S_1 having the reasons that S_2 has for knowing *p*. Hardwig "unpacks" these conditions in the following way: "(1) S_1 knows that S_2 says *p*; (2) S_1 believes (and has good reasons to believe) that S_2 is speaking truthfully, *i.e.*, that S_2 is saying what she believes; (3) S_1 believes (and has good reasons to believe) that S_2 (unlike S_1) is in a position, first, to know what would be good reasons to believe *p* and, second, to have the needed reasons; (4) S_1 believes (and has good reasons to believe) that S_2 actually has good reasons for believing *p* when she thinks she does"[7].

In the present paper, I do not argue for Hardwig's proposal for a necessary condition for knowing *p*, even though it is important to see what follows if his proposal is rejected, and if, indeed, there are no adequate proposals for a social conception of knowing *p*[ix]. The consequence is that either there would be no knowledge where, intuitively, we take it to be the case that there is knowledge or that knowledge "that *p*" would not be a relation between an individual and *p*, but rather a relation between many individuals and *p*. The latter is a distinctively Deweyan conception of knowledge; however, if we accept this account, it would be disastrous for any account of how privacy and trust can cohere in e-networks[x]. For if n individuals could be said to know that *p*, where no single individual knows that *p*, then there would be occasions in which private information is known by the entire network of agents, even though there are safeguards to prevent this from happening. Consider the following example: Suppose that private information about an individual is decomposed into m strings and that no agent is ever given access to more than m-k strings and that no agents can communicate with one another. It would appear that this is a secure system and that there could be no breach of privacy. However, if we adopt a Deweyan conception of knowledge, then the entire community of agents—both human and artificial—knows *p*. This would mean that privacy would be violated where there is at least one agent who is prohibited from knowing *p* because that agent is taken to be a security risk. Moreover, it is clear that if there are agents outside the local system with whom agents in the system interact, then those agents might also know *p* as well. Indeed, there could be a set of conditions under which all agents in the world know *p*. This is a serious problem, though it arises only on a Deweyan conception of knowledge.

Of course, if there were conclusive arguments that a Deweyan conception of knowledge is incoherent, then the worry just sketched in the previous paragraph would evaporate. But there are no conclusive arguments that Deweyan conceptions of knowledge are incoherent. The problem, then, is the following: So long as there are undefeated accounts of knowledge under which all agents know *p* and where some of those agents are security risks who should be prohibited from knowing *p*, any account of how trust and privacy can cohere in an e-network will be seriously compromised. The compromise occurs when an agent believes that she cannot trust other agents because it might be the case that all agents know *p*, since it might be the case that a Deweyan conception of knowledge is true. Clearly, we cannot simply adopt some conception of knowledge that is congenial to our purposes in the scenario of trust and privacy in e-networks. There will be, in the total scheme of things, accounts of knowledge that are correct and accounts that are not correct; but where we do not know which accounts are correct and which are not, we cannot rule out accounts under which bad consequences occur. It is a bad consequence that every agent in an e-network (and possibly outside the e-network) knows *p* in the context of an account of trust and privacy, since we could not trust that the e-network protects privacy, since we could not trust each agent to protect privacy (by either trusting that the agent not access private information or trusting that agents who do access private information not share it with agents who are not allowed to access it).

Indeed, things are even worse than that—since the very idea that every agent knows *p* leads one to think that there could be no protection of privacy, no matter what actions one performs or what safeguards and/or directives are in place. Where an agent thinks that privacy cannot be protected, trust relations that would protect privacy cannot be established. That is the problem for any account of trust and privacy in e-networks. Until the final scheme of things has been established in philosophy, or until a Deweyan conception of knowledge has been conclusively refuted, it is open that such an account could be viable. If so, it would be difficult to establish a relation of trust between agents in e-networks that is designed to protect privacy. There is certainly an affinity between this problem and the problem of skepticism in epistemology. But the difference is interesting—only a philosopher who is perhaps mentally ill would seriously deny the everyday knowledge claims that one makes daily; but where there are open various accounts of knowledge on one of which trust is impossible, it follows that trust is impossible no matter what account one adopts. As long as it is possible that all agents in an e-network know *p*, trust cannot be established. Since this problem appears to be irresolvable in the absence of conclusively refuting a Deweyan conception of knowledge, I will simply assume, in the remainder of this paper, that such a conception of knowledge is not a viable option. (If the skeptical problem is not considered by any of the agents in an e-network—or, indeed, in any social interaction—(because no one agent knows of it), then it will not arise. This is, however, a sociological matter.)

However, one final word on the problem. In his recent book, *Privacy*, Raymond Wacks points out that there are situations in which, although there is not an actual mechanism in place for invading one's privacy, the mere idea that there might be such a mechanism in place is sufficient to constitute an invasion of privacy[8]. He says: "My actions have not been monitored, yet subjectively my equanimity has been disturbed. The mere presence of a device that appears to be observing and recording my behavior (even though it is not doing that—J. B.) is surely tantamount to the reality of my unease. In other words, it is the *belief* that I am being watched that is my grievance. It is immaterial whether I am in fact the subject of surveillance. My objection is therefore not that I am being observed—for I am not—but the possibility that I may be"[xi] [8]. In such cases, counterfactuals of the form "If the device had been recording, there would have been an invasion of privacy, but the devices were not recording" do not deflect the force of the objection to their presence, since it is the idea that the device could invade my privacy which is, itself, the invasion of my privacy. Thinking that I am being observed, even though I am not, it is sufficient to make me feel quite uneasy.

We can easily extract from the essential features of the case enough to construct a parallel case in an e-network. A human agent will not feel that the data about him in an e-network is protected from those agents who want to acquire it, illegally, if she knows that the algorithms for security matters are badly designed. Surely, in such a case, we would say that the agent may feel an invasion of her privacy, even if the algorithms perform perfectly well whenever they are deployed. It is the idea that the data about her might be illegally expropriated which constitutes the invasion of her privacy. This, of course, invites the "princess and the pea under the mattress" objection: If an agent simply believes that the data about her might be illegally expropriated, that cannot count as an invasion of privacy, since simply thinking that thought is not enough. One can think that thought for different reasons, and some reasons are not sufficient to warrant that there has been an invasion of privacy. If I think that thought because someone has bopped me on the head, it is not enough to count as an invasion of privacy. Clearly, the reason for thinking that thought has to warrant the claim that there has been an invasion of privacy. In the cases above in which a Deweyan conception of knowledge would make it the case that all agents in the e-network know *p* (where *p* is private information about the agent in question), it is enough to warrant the charge of invasion of privacy. However, there is a grey area in which there will be cases that can fall one way or the other.

4. The Union/Intersection Condition and Affiliated Problems for It

Social conceptions of knowledge differ from individualistic conceptions in a distinct way: On the latter account of knowledge, a single agent knows *p* when she believes *p*, is justified in believing *p* (discounting for "Getteried" beliefs), and *p* is true. On a social conception of knowledge, a single agent can know *p*, without being justified in believing *p* in the same way that an agent is justified on the individualistic account. Agents in social accounts do not have to possess the justification conditions for *p*. That is, on a social conception of knowledge, for certain propositions *p* (that cannot be known on an individualistic conception of knowledge, but only on a social conception of knowledge), there will be n agents, each of whom knows some proposition p_n and such that the proposition *p* is the union of p_n. Moreover, it is the case that the intersection of any two agents with respect to what they know in common is the empty set. That is, the intersection of $K_{\text{agent}(i)}\ p_i$ and $K_{\text{agent}(j)}\ p_j$ is the empty set. The role of trust in this situation is that each agent trusts the other agent not to broadcast information that is sent to them from another agent. (Once an agent sends information to another agent, the information is "cleanly" transferred and the sending agent no longer has access to that information, while the agent who has received that information does have access to it. Clearly, this is a simplification. There are other schemes in which agents who send information retain access to it. This will complicate matters, and will produce a situation in which the intersection condition on a social conception of knowledge is no longer satisfied. Certainly, social conceptions of knowledge allow all sorts of conditions—the intersection condition is just one of many such conditions.)

One problem that arises for the union/intersection condition that must be addressed is that of collusion under a guise. This is an interesting problem that arises because there are two different kinds of trust—epistemic and moral. An agent might misrepresent to another agent that it needs information in order to—eventually—know that *p*. The agent that is asked to send that information might reason: "The other agent needs this in order to establish *p* and I have reason to believe that this agent is reliable and satisfied the four conditions on a social conception of knowledge; thus I know what the other agent knows (in order to know *p*)". The agent sends the information to the first agent. But the first agent has falsely represented that she needs that information for epistemic reasons. Indeed, she does need the information—but the reason why she does is that she wishes to violate the privacy of the individual that information concerns. Where agents do not take into consideration that they can be trusted in one sense, but not trusted in another sense, opportunities for violation of privacy are available.

It is necessary, then, to design a system in which there are protections that take into account that not all agents will know that there are (at least) two different kinds of trust relations in play, unless the system can be designed, in the first place, to ensure that all agents know there are (at least) two different kinds of trust relations in play and that they can reason about the interactions between the two different kinds of trust. The point is that, just as there can be breaches of moral trust in the form of cunning and misrepresentation, so, too, there can be the same kinds of breaches for epistemic trust, as well as guise problems where one form is trust is misrepresented as another form of trust (that is, the latter arises only where there are two—or more—distinct forms of the trust relation). It would be a mistake to design a system in which such a possibility is not taken seriously.

5. Are There Necessary Conditions for a Moral Trust Relation Grounded on an Epistemological Analysis of an E-network?

A different set of cases arise in situations in which an agent reasons that another agent will reason explicitly about whether the epistemological conditions necessary for a moral trust relation in an e-network have been satisfied. Thus, we will need to examine the question of whether there are epistemological conditions that ground the necessary conditions for a moral trust relation in an e-network. (Clearly, by the necessary conditions for a moral trust relation in an e-network, we mean how the standard necessary conditions for a moral trust relation are filled in relative to an e-network. On this view of the necessary conditions for a moral trust relation, they are parameterized and the parameters are filled in, depending upon the context.) What will complicate the epistemological analysis is the kinds of interactions between humans and artificial agents. We should distinguish then, four kinds of trust relations: (i) where a human agent trusts a human agent, (ii) where a human agent trusts an artificial agent, (iii) where an artificial agent trusts a human agent, and (iv) where an artificial agent trusts an artificial agent.

If it were the case that human and artificial agents were indistinguishable cognitively and morally, then this typology would not be necessary. However, given that there are sufficiently many differences between human and artificial agents, there will be differences in the conditions for trust relations to take root. For example, consider an epistemic trust relation. Given an artificial agent with a reliable capacity for computing integrals (of a certain kind), one would epistemically trust that agent, but not epistemically trust a human agent whose capacity for computing integrals is second-rate at best. In the case of moral trust relations, one might morally trust a human agent because of the evidence of good will that agent exhibits, while not trust an artificial agent, because the latter does not have the capacity to exhibit good will. Further, we can also distinguish between those agents, human and artificial, for whom there is good reason to think they will not broadcast private information and those for whom there is not a good reason to think they will not broadcast private information.

This distinction will lead to an 8 cell matrix of trust relations. If we also make these distinctions at the level of the entrustor—that is, we distinguish entrustors in terms of those for whom there is a good reason to think they will not broadcast and those for whom there is no good reason to think that—then there will be a 16 cell matrix of trust relations. I don't propose that the matrices actually get filled in (at least, here), since it would take too much space. The point of bringing up these distinctions is that any agent—human or artificial—that reasons about the epistemological conditions present in an e-network in which there is protected data and in which the issue of privacy is salient may well reason that there are so many possibilities to consider that an explicit consideration of each of them is computationally infeasible. In which case, any agent will not be able to reason about them, and thus will not be in a position to see if the epistemological conditions necessary for a moral trust relation are in place. Given that this cannot be solved computationally, such agents will conclude that there cannot be a moral trust relation. The problem, then, is to specify the conditions under which such an inference will not be valid, and to ensure that those conditions are satisfied in e-networks (where privacy considerations are salient) and that agents working in those networks know that they have been satisfied.

6. The Deception Problem

Although agent A might not have any expertise concerning knowledge claim *p*, if agent B does and agent A epistemically trusts B, then A can know *p* without having any reasons intrinsic to *p* to justify that belief. Rather, justification goes through B. A has a good reason to believe *p* because she epistemically trusts B and B has a good reason to believe *p*. However, it has been noted that where B deceives himself concerning the extent of his knowledge or its reliability, then A does not have good reasons to believe *p*, and, in fact, is not justified in epistemically trusting B [9]. Certainly, this seems to be reasonable, for if B is prone to deceiving himself about how much he knows, then A is epistemically affected as well, since the credibility of B is diminished. Suppose that there are n agents, all of whom have specific knowledge claims p_1, p_2, p_3, … , p_n, none of whom intrinsically know the knowledge claims of any other agent, and where knowledge claim *p* is the union of p_1, p_2, p_3, … , p_n. However, if agent a_i deceives himself into believing he knows more than he does, then, since a_i is part of the knowledge claim *p*, A cannot trust a_i. If so, then a_i does not know *p*. But notice that this is also true of any other agent in the knowledge network as well. That is, for any agent a_j, where $i \neq j$, a_j does not know *p*. Since a_i trusts a_j, then even a_i does not know *p*.

So it would seem that what initially appeared to be a reasonable restriction on social knowing now appears to be much too strong. Wherever anyone in a knowledge network deceives himself into believing he knows more than he does, then no one, whether in the knowledge network or not, knows anything socially. Social knowing, then, would be quite hard to attain. Notice that this restriction would not be acceptable for individual knowing. That is, suppose that A believes *p*, is justified in believing *p*, and *p* is true. Discounting Gettier cases, we say that A knows *p*. Suppose that A deceives herself into believing that she knows *q*, where *p* and *q* are independent propositions. Such deception would not show that A does not know *p*. It would only show that A does not know *q*. (If the justification for *p* involved *q* in some way, then it would be the case that A does not know *p*, but that is because it would be true that A is not justified in believing *p*, since A has no justification for his belief that *q*.)

It follows that such *a* restriction would be catastrophic for e-networks involving perhaps hundreds of thousands of artificial agents. If the epistemic trust relation depends upon what another agent knows or might come to know, then *a* rampant skepticism infects the e-network, since it is always possible that there is some agent who falsely believes he knows more than he does. Indeed, an agent could easily come to falsely believe she knows more than she does in any number of different and quite natural ways. Here is one such way: A knows *p*, because it has been routed to A from some agent B and A infers, on the basis of this, that A also knows *q*, where *q* is the proposition that *p* has been routed to the agent from agent B. Actually, it was agent C who routed *p* to him. Unless each agent keeps track of the routing of information to him, for all packages of information and for all agents in the e-network, an agent might easily come to have false beliefs about what it knows. In which case, not only does the agent not know *q*, it does not know *p* either, since knowledge of *p* depends upon whether other agents know it as well.

Once again, social knowing is vulnerable to *a* powerful skepticism where the restriction that engenders the deception problem—that an agent need deceive itself as to what it knows—is in place. Individual knowing does not need such *a* restriction, but the problem with individual knowing in an e-network is that it is computationally infeasible. That is, for any agent a_i (whether artificial or human), a_i cannot know *p* (for any *p*) because there are too many pieces in the knowledge network (for *p*) for a_i to know. Social knowing is a way of avoiding computational intractability. But the deception problem creates a trilemma: Either there is no social knowing of *p*, for any *p*, (because of the skepticism problem it engenders) or else all knowledge that *p* must be individual knowledge (in which case, no agent knows *p*, for any *p*, since to individually know any *p* an agent must possess computational resources he does not have and cannot have), or there is Deweyan knowing, in which case all agents as-a-whole know *p* (though no single agent, or subgroup, in the whole knows *p*). Unless the restriction is relaxed in some way, then there are situations in which no one can know *p*, since to know *p* socially engenders the deception problem, to know *p* individually requires computational resources no individual possesses, and to Dewey-know *p* is to know *p* only in a trivial sense.

7. The Deception Problem, Social Knowing, and *de re* and *de dicto* Propositional Attitudes

Where *p* is information about one's personal life, knowledge that *p* is not merely a matter of having a good reason to believe that *p* is true. In addition, knowledge that *p* may require an interpretation of *p*. Suppose that *p* is a social security number. Knowing that it is a genuine social security number may require associating it with someone's name. If there is no possibility of associating the social security number with a name, then can an agent be said to know that it is a social security number? One response to the preceding is to deny that an agent does not know it is a social security number, since that must be true, given that all 9-digit numbers entering the network are social security numbers. Rather, what the agent does not know is whose social security number it is. Thus, a distinction is made between the proposition that the 9-digit number is a social security number and the proposition that the 9-digit number is, say, Jack's social security number. Unless the 9-digit number is properly interpreted to be Jack's social security number, it is not personal knowledge about Jack. One can envisage a situation in which a list of social security numbers is known by one agent and a list of names is known by another agent, and there is a 1-1 map of names into social security numbers and another 1-1 map of social security numbers into names. But only if each agent shares its information with the other agent can each agent know that it is, for instance, Jack's social security number. Now suppose that names are broken into pieces, and that social security numbers are broken into pieces. This, in turn, creates n lists, where there are m lists for social security numbers and l lists for names. Each list is known by an individual agent, though no agent knows more than one list. This kind of scenario might be implemented in e-networks to respect privacy of personal information.

Now if agent A knows that agent B knows *p*, can we say that agent A knows *p* as well? Consider the case where A knows of each item on the list of social security numbers that it is a genuine social security number and B knows of each item on the list of names that it is a genuine name. If A knows B knows this and B knows A knows this, then on a social conception of knowledge, A also knows what B knows. This sounds odd—it appears that a social conception of knowing is not appropriate in this context. There is a well-known distinction between *de re* and *de dicto* propositional attitudes. For example, 'John believes there are spies' is ambiguous: It can mean either (i) there are certain persons whom John believes to be spies or (ii) John believes there are people who are spies, but does not have any particular person in mind.

Some have argued that *de re* attitudes reduce to *de dicto* attitudes, though this would mean that anyone who believes that the tallest spy is a spy will also believe, by existential generalization, that there is someone who is a spy[xii] [10]. Since propositional attitudes include wanting and knowing, there will be similar problems for these attitudes as well. In the context of trust and privacy in e-networks, then, there is the following kind of problem. Take, for instance, "wants". Suppose an agent wants data, but no particular data, while another agent wants particular data. (This is analogous to the distinction between someone wanting to marry a Scandanavian and someone wanting to marry a particular Scandanavian.)

Suppose that there is an agent who wants particular personal data (and thus will violate privacy of the individual involved), but does not want any other personal data and that there is another agent who wants personal data (but not any particular personal data) in order to violate privacy, say, just for the sake of violating privacy. Now someone who thinks that *de re* attitudes reduce to *de dicto* attitudes—call them S_1—will believe that the first agent should not be given any data at all, since that agent will want any data, and not just some particular data. On the other hand, someone who believes that *de re* attitudes do not reduce to *de dicto* attitudes—call them S_2—will think that the first agent wants particular personal data to violate privacy, but does not want any other personal data in order to violate privacy. Thus S_2 will think that it is acceptable for the first agent to be given personal data other than the personal data that agent wants for the purpose of violating privacy, while S_1 will think that it is not acceptable to give the first agent that personal data, since S_1 believes that the first agent will violate trust and invade privacy for all personal data it acquires.

There will be a conflict between S_1 and S_2. It is clear that S_1 will not trust S_2 and that S_2 will think that S_1 is making demands that are not justifiable. Thus, in the eyes of S_2, S_1 has lost epistemic credence, while in the eyes of S_1, S_2 has lost normative responsibility. Where an agent has lost epistemic credence, another agent will not be able to establish an epistemic trust relation with that agent. And where an agent has lost normative responsibility, another agent will not be able to establish a moral trust relation with that agent. Here the fact that there are disputes over whether *de re* propositional attitudes reduce to *de dicto* propositional attitudes generates failure of trust relations of both kinds—epistemic and moral, though in differing agents and in differing ways. However, there can also be cases in which a single agent can be accorded neither moral nor epistemic trust.

These problems also challenge a social conception of knowing, since the propositional attitude "to know" can generate the same kinds of problems that the propositional attitude "to want" generates. (Moreover, they also haunt an individualistic conception of knowing, for the same reasons.) The way in which they challenge that conception of knowing is certainly more interesting (and more complex) than that of individualistic knowing. Although there is not enough space to describe the specific conditions under which this happens in social conceptions of knowing, some results are, for instance, that in the case of social knowing, someone who thinks that *de re* propositional attitudes reduce to *de dicto* propositional attitudes might think that all agents in an e-network know *p*, and thus, where there is an agent in the e-network who wants a particular piece of data (to violate privacy), that not only does that agent want all data to violate privacy, but that all agents in the network cannot be trusted since they might give any particular piece of data to that agent. Thus the e-network will suffer gridlock, and will shut down. There are other kinds of problems, as well. For instance, under what conditions do you know who someone is? Simply being told their name does not mean that you know who they are. The conditions under which knowledge that *p* is genuine knowledge of *p* can, in some situations, be quite slippery, and relatively contentious. Thus agents might have personal information, but no knowledge, where they have not been able to properly interpret the data. But where there is proper interpretation, there is also knowledge.

8. The Deception Problem and Computational Intractability

Knowledge of data is a central problem of privacy in e-networks, but there are several problems that must be addressed before reasonable principles concerning access to data can be formulated. An agent that simply acquires information does not, ipso facto, have knowledge, even if the mechanism by which acquisition occurs is reliable (externally) or justified (internally). For information to have the status of knowledge, the information must be interpreted in a certain way. For instance, if an agent acquires a social security number, that information is not knowledge unless the agent knows that it is a social security number. That is, the agent must have the capacity to interpret the number as a social security number.

There are various kinds of trust relations that occur in e-networks, some of which are in considerable tension with one another. An example of such tension is that between (i) a trust relation in which agent A trusts that agent B is reliable with respect to the data which B routes to A and (ii) a trust relation in which A trusts that B will not broadcast that data—and violate the privacy of whomever the data is about. Similarly, B trusts that A will not broadcast the data and that any data routed from A to B is reliable. Where there are a great many artificial agents in an e-network, all of which might have access to any given piece of data, certain structural problems arise. For instance, are (i) and (ii) unequally distributed among different artificial agents? That is, are there cases in which, for example, A trusts B in sense (i), but not in sense (ii), while A trusts C in both sense (i) and sense (ii)? If so, how an artificial agent can recognize that such an arrangement is necessary is itself problematic, since it appears that computing such arrangements will be highly infeasible.

A simple argument that it is computationally infeasible is that, given N artificial agents, M properties (taken as necessary conditions for securing trust relations) and K trust relations, there are N × M × K possible arrangements, each of which has to be examined separately. However, the examination of any particular arrangement will itself require computational resources, since the agent performing the examination will have to consider the different kinds of circumstances in which it might be involved, as well as the kinds of punishments and rewards that are available for different kinds of arrangements in different kinds of circumstances.

9. Defaulting, Reputation, and Resource Conservation in E-networks

Without an agent's reputation at stake, without punishments and without moral structures (of some kind), there would be no reason not to default—to violate the trust relation—in cases where the agent will profit from such a default. If the monetary stakes are sufficiently high, an agent might violate privacy by breaking trust. This kind of case is different from the previous cases, in which there is a conflict between two kinds of trust—epistemic and moral—and the result is that privacy is violated when epistemic trust is sustained at the expense of breaking moral trust. Will an agent's sense of its reputation serve to eliminate or severely limit defaulting behavior that results in violations of privacy? Experiments involving artificial agents in e-networks done by Barbara Grosz and her collaborators, in a different context—that of coalitional behavior involving artificial agents—show that where a sense of reputation is taken to be important by the artificial agent, defaults from cooperation in a coalition will be limited. [11] That raises the question—what should trusting artificial agents think of the value of reputation? Where epistemic trust is in play, reputation will be different from where moral trust is in play. For instance, an agent who is considered to be reliable with respect to knowledge deliverances will value her reputation as one who is valued in this respect. But this kind of valuation is certainly different from valuations in cases of moral trust, where an artificial agent will value having performed the morally appropriate action.

The resource conservation that results from trust relations—whether epistemic or moral trust relations—is certainly a benefit, just as an agent's reputation is a benefit (in those situations where reputation provides that agent with opportunities that agents who lack reputation would not have). But in situations in which there is a large profit from defaulting, the question arises as to when defaulting occurs. That is, what specific values must reputation and resource conservation have before defaulting is a viable option for an agent? When resource conservation benefits outweigh the profit from defaulting, it is clear that defaulting will not occur, and, moreover, it suggests that there is no need for reputation. This is an interesting result, since in the systems that Grosz and her collaborators have studied, and in which they have performed numerous experiments, it is the measure of social consciousness (one's reputation) that impedes defaulting. [11,12] However, there are various factors that go into reputation, and that also are necessary for establishing both moral and epistemic trust relations. These factors will complicate the analysis of the conditions under which resource conservation and reputation outweigh defaulting. Even in the case of epistemic trust, there are many factors that are necessary for establishing epistemic trust, such as epistemic competence (of the agent who knows *p*), independence, and self-reliance. [7]

Certainly, an agent—whether human or artificial—who sees herself as part of a team engaged in the pursuit of knowledge, or engaged in the pursuit of some economic goal, will be less inclined to default than an agent who sees herself as independent and self-reliant. This, however, creates a problem of the following kind. Independence and self-reliance are typically not necessary when a trust relation is in place, though they are necessary for establishing, in the first place, a trust relation. In many cases, the nature of resource conservation consists in filling in for the kinds of results that are brought about by independence and self-reliance. If an agent shows that he is not self-reliant or shows that he is not independent (for example, he shows that he depends upon monetary compensation from another agent), he will be seen in the eyes of someone who is thinking of whether to trust him or not as perhaps not worthy of trust. Once, however, the trust relation has been established, there is no longer a need for the trusted agent to exhibit independence and self-reliance, since that agent may help himself to the benefits of the trust relation. There are various benefits, though the ones we have described in this paper have to do with conservation of computational resources. It is in virtue of the trust relation—which has the form of a dependency between two agents (the trustor and the entrusted) that the benefits accrue to each agent.

However, the state-of-affairs described in the preceding paragraph raises an issue which, prima facie, appears paradoxical: A property is necessary for establishing trust, which then renders that property superfluous. But the air of paradox vanishes when trust is viewed as a dynamic process, not as a static one. That is, once the conditions that are needed to establish a trust relation have done their work, they are no longer needed to sustain the trust relation. But the sense of reputation that an agent entertains will change as the agent's sense of their own independence and self-reliance changes. If the revised sense of reputation is commensurate with the original sense of reputation, then the level of defaulting will remain the same. However, where the sense of reputation is not commensurate, the level of defaulting will change—it will go higher where reputation diminishes and lower where it is elevated.

Now consider the following situation—that of transmission of private information in public institutions. This kind of transmission is not avoidable and is not a reason to think that an agent has abrogated their claim to protect privacy, where that agent is entrusted with private information, as well as with public information (not necessarily about the same individual that the private information is about). This kind of situation is different from a situation where an agent intentionally performs a private act in public, and information about that activity is transmitted throughout the public institution. Now consider an agent who has been entrusted with protecting the privacy of information about a certain individual. The agent has been shown to be both epistemically and morally trustworthy by exhibiting the virtues of independence and self-reliance (perhaps in different ways for each kind of trust). Once entrusted, that agent no longer need exhibit those virtues, at least with respect to acquiring the resource conservation benefits of the epistemic trust relation and those of the moral trust relation.

However, this creates a problem in the sense that once the agent benefits from being entrusted, his sense of reputation will change, in that the features that partially defined his original sense of reputation are no longer present. But this same agent now benefits from conserving computational resources and thus is receiving benefits even though her sense of reputation has diminished. That agent can easily infer that diminished reputation can result in his receiving benefits and thus, that defaulting may be the next best thing to do. Of course, this situation could be avoided if the agent were designed so that it either did not have the features of independence and self-reliance—in which case, it would not be deemed trustworthy, or did not assess those features in establishing its reputation—in which case, it would not be able to have a sense of reputation that was appropriate for the job of protecting privacy in public institutions, such as public e-networks.

10. Concluding Comments

The extent to which an agent can trust another agent where privacy issues are raised is a new and important problem in the context of e-networks. A feature of trust that has not been recognized in the literature—that it conserves valuable computational resources—is Janus-faced. The resources saved in epistemic and moral trust contexts may well come at the expense of compromising, if not losing, privacy. On the other hand, privacy could not be enforced unless resources are conserved via a trust relation. There are also subtle interactions between epistemic and moral notions of trust where privacy issues are raised in the context of e-networks.

There are several hard problems raised in analyzing the connections between trust, privacy, and resource conservation issues: Formulating a social conception of knowing, the Union/Intersection condition, epistemological analyses of e-networks, the deception problem, social knowing and *de re* and *de dicto* propositional attitudes, computational intractability, and the relation between defaulting and reputation. These problems are surprisingly robust; their solution requires interdisciplinary ventures and is only a first step toward a comprehensive theory of privacy and trust in e-networks.

Acknowledgements

This is the first of two papers on the subject. I wish to thank Herman Tavani for insightful comments and supererogatory editorial work.

References and Notes

1. Hersh, S. The Online Threat—Should We Be Worried About a Cyber War? *New Yorker* **2010**, 44-55.
2. Baier, A. Trust and Anti-Trust. *Ethics* **1986**, *96*, 231-260.
3. Walker, M. *Moral Repair*; Cambridge University Press: New York, NY, USA, 2006.
4. Govier, T. *Trust and Human Communities*; McGill-Queen’s University Press: Montreal, Canada, 1997.
5. Jones, K. Trust as an Affective Attitude. *Ethics* **1996**, *107*, 4-25.
6. Baker, J. Trust and Rationality. *Pac. Phil. Q.* **1987**, *68*, 1-13.
7. Hardwig, J. The Role of Trust in Knowledge. *Pac. Phil. Q.* **1991**, *88*, 693-708.
8. Wacks, R. *Privacy*; Oxford University Press: New York, NY, USA, 2010.
9. Ritchie, B. Epistemic Akrasia and Rational Requirements. In *Proceedings of the 2010 Northern New England Philosophical Association Conference*, St. Anselm.s College, 16-17 October, 2010, Birkenhead, UK.

10. Kripke, S. *Unrestricted Exportation and Some Morals for the Philosophy of Language in Collected Works*; Oxford University Press: New York, NY, USA, 2011; Volume 1.
11. Grosz, B.; Kraus, S.; Sullivan, D.; Das, S. The influence of social norms and social consciousness on intention reconciliation. *Artif. Intell.* **2002**, *142*, 147-177.
12. Buechner, J.; Tavani, H. Trust and Multi-Agent Systems: Applying the "Diffuse, Default Model" of Trust to Experiments Involving Artificial Agents. *Ethics Inform. Tech.* **2011**, in press.
13. McCarthy, J.; Hayes, P. *Some Philosophical Problems from the Standpoint of Artificial Intelligence in Machine Intelligence 4*; Meltzer, B., Mitchie, D., Eds.; Edinburgh University Press: Edinburgh, UK, 1969.
14. Ginsberg, M. *Readings in Nonmonotonic Reasoning*; Morgan Kaufman Publishers: Los Altos, CA, USA, 1987.
15. Gettier, E. Is Justified True Belief Knowledge? *Analysis* **1963**, *23*, 121-123.
16. Dewey, J. *The Quest for Certainty*: *A Study of the Relation of Knowledge and Action*; George Allen and Unwin: London, UK, 1930.
17. Bentham, J. Panopticon. In *The Panopticon Writings*; M. Bozovic, Ed.; Verso: London, UK, 1995.
18. Buechner, J. Trust as a Means of Conserving Valuable Computational Resources. Unpublished work, 2010.

[i] Hersh [1] notes that "the federal government currently spends between six and seven billion dollars annually for unclassified cyber-security work, and, it is estimated, an equal amount on the classified portion." (p. 48)

[ii] In these cases, the reliability of an algorithm used to assign information to various artificial agents is what matters in deciding whether to trust that an e-network does not violate privacy. In which case, a human agent might (or might not) trust the designer of the algorithm (with or without knowing who she is, as well). There are, at least, two different kinds of trust relations in Web 2.0. One relation is to trust other people or agents with personal information and the other relation is to trust other people or agents to not access personal information about oneself. In the first kind of trust relation, agents are entrusted with personal information, while in the second kind of trust relation, agents are entrusted with not acquiring personal information. Under what general conditions do both kinds of relations of trust obtain? Of course, there can also be hybrid relations of both kinds. For instance, an agent can trust another agent with personal information A, while trusting that same agent not to acquire personal information B. This naturally complicates the relationship, especially when there are multiple agents in complex networks, and multiple kinds of personal information, some of which may be deducible or inferable from other kinds.

[iii] There is a large literature on frame problems. The first paper in the field is the classic one from John McCarthy and Patrick J. Hayes [13]. For a good selection of important articles on solutions to the frame problem, see M. Ginsberg (ed.) [14].

[iv] I develop and defend the idea that a trust relation provides a means of conserving valuable computational resources (and other kinds of resources as well) in "Trust as a Means of Conserving Valuable Resources," (unpublished manuscript).

[v] For a discussion of the normative notions that underlay the notion of trust, see [12].

[vi] For a good discussion of epistemic akrasia, see [9].

[vii] "Gettiered" beliefs are beliefs that are true and justified, but do not count as knowledge [15]. How to accommodate Gettiered beliefs in a theory of knowledge is an open and vexing problem.

[viii] It is clear that there is a need for a careful discussion of knowing, believing, and accessing. There are, of course, well-known differences between believing and knowing, but there is less known about the differences between accessing and believing and between accessing and knowing. Certainly, these distinctions need to be made where algorithms determine which agents acquire information, as well as control the flow of information through the network. If an agent accesses *p*, it does not follow that she believes *p*, since she might take p to be false. An agent might not even take *p* to be about anything, *i.e.*, simply taking *p* autonomously (to be, for example, a set of symbols, period). On the other hand, if accessing of information occurs in a veridical environment and an agent knows this to be so, then whenever an agent accesses p she has a reason to believe *p*—namely, that p is true. Moreover, if the agent knows that the environment is veridical, then she is justified in believing that p is true. Thus an agent knows *p* given that she has accessed *p*. Of course, social and business e-networks are not veridical. Information that is true at time t_l might no longer be true at time t_2. Moreover, it might become true once again at time t_3. Thus, information that p can become true (if previously false) and become false (if previously true) at any given time in such networks.

Attention needs to be paid to these distinctions in designing e-environments in which privacy issues are salient, since there will be cases in which accessing p ipso facto gives one knowledge that *p*, or believing that *p* gives one knowledge that *p*. Moreover, an agent might know *p* without having accessed *p* (especially on a social conception of knowing). Obviously, the complexity of the environment depends upon how many agents there are and on both how much information flows through it and the kinds of information in the information flow.

[ix] There is an independent way of motivating a social conception of knowing differently from the epistemic motivation. Under what conditions does the trust relation flourish? Where A knows that B can do what A cannot do, and B knows this, and B feels responsible for doing what A cannot, and A knows this, conditions are in place for the trust relation. These conditions do not define the trust relation. Certainly, there can be a trust relation between A and B even when B does not know that A cannot do what B does, and B might even believe that A can do it. Suppose that A is both independent and self-reliant. However, she encounters a situation in which she cannot do something she needs to do to achieve her goals. B knows this, and also knows she is independent and self-reliant. B takes these two virtues that A possesses to be reasons for helping A—that is, she helps A do what A cannot do because she believes that A would do it if she could (being independent and self-reliant), but since she cannot, she genuinely needs help. Now suppose that the situation in which A requires help to achieve her goals becomes standard (or typical) in the sense that everything that she does occurs in that kind of situation. Thus, all of her goals can be achieved only if there is another agent who can help A achieve them. But, in that case, B might no longer have reasons to help A, since in that situation A is no longer independent and self-reliant. Of course, if the situation were not like that, then A would be, *ceteris paribus*, independent and self-reliant. But sensitivity to such counterfactuals does not, *ipso facto*, provide B with reasons for helping A. (At least, not in the same way that it did when such kinds of

situations rarely occurred.) This impasse can be bypassed where one adopts a social conception of knowing and abandons an individual conception of knowing. In social conceptions of knowing, a relation of trust between two knowers replaces an individual relation between a knower of *p* and evidence that *p* is true.

[x] The Deweyian conception of knowledge is the view that some propositions can only be known by groups of persons, but not by individual persons. Although some have argued that Dewey did not hold this view, most Dewey scholars take it that Dewey did hold the view. As for Dewey's views on the subject, see [16].

[xi] The reader familiar with Jeremy Bentham's "Panopticon" will see that Bentham clearly anticipated the point made by Weeks [17].

[xii] There is an enormous literature in the philosophy of language on this distinction, but perhaps the best discussion of it occurs in [18].