MANAGEMENT SCIENCE
Vol. 57, No. 2, February 2011, pp. 393–412
ISSN 0025-1909 | EISSN 1526-5501 | 11 | 5702 | 0393

informs

DOI 10.1287/mnsc.1100.1279

"I Think You Think I Think You're Lying": The Interactive Epistemology of Trust in Social Networks

Mihnea C. Moldoveanu
Desautels Centre for Intregrative Thinking, Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada, micamo@rotman.utoronto.ca

Joel A. C. Baum
Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada, baum@rotman.utoronto.ca

We investigate the epistemology of trust in social networks. We posit trust as a special epistemic state that depends on actors' beliefs about each others' beliefs as well as about states of the world. It offers new ideas and tools for representing the core elements of trust both within dyads and larger groups and presents an approach that makes trust measurable in a noncircular and predictive, rather than merely postdictive, fashion. After advancing arguments for the importance of interactive belief systems to the successful coordination of behavior, we tune our investigation of trust by focusing on beliefs that are important to mobilization and coordination and show how trust functions to influence social capital arising from network structure. We present empirical evidence corroborating the importance of higher-order beliefs to understanding trust and the interactive analysis of trust to the likelihood of successful coordination.

Key words: organizational studies; strategy; networks graphs; theory; design; information; philosophy of modeling
History: Received October 10, 2007; accepted September 28, 2010, by Jesper Sørensen, organizations and social networks.

Introduction: Trust as an Interactive Epistemic State

In this paper, we model the systems of knowledge and belief involved in the phenomenology of trust in social networks. We argue that trust can be understood as a system of interactive beliefs that the trustful and the trusted share about each other's propensity to register and report relevant information to one another. Prototypical of such relevant information are propositions that represent the backbone of coordinative, communicative, and collaborative activities: What matters to coordination and mobilization in a wide range of contexts is coherence within what actors think about what other actors think (level 2 beliefs) and what actors think others think they think (level 3 beliefs) about a set of issues that are important to the interaction. We argue that trust in competence and integrity safeguards coordination and cooperation among networked agents, and therefore that trust in competence and integrity should be correlated with coherent hierarchies of higher-level beliefs, which are required for successful coordination and mobilization. We marshal evidence that is consistent with this argument.

After first motivating our approach to representing trust, we build a language for describing knowledge states of actors in a social network (their "epistemic states") and show how trust can be defined as an epistemic structure that depends on actors' second- and third-order beliefs. The description language enables us to (a) represent the knowledge states of actors within a network, (b) demonstrate the importance of actors' higher-level epistemic states to the achievement of successful coordinative action within the network, and (c) analyze trust as a modal epistemic state, wherein actors have coherent beliefs about other actors' proclivity to know the truth (trust in competence) and to speak the truth (trust in integrity). We posit that actors who trust one another will be more likely to share coherent interactive belief hierarchies ("what I think matches what you think I think"), and show how trust structures form the backbone of epistemically coherent subnetworks of agents. We use this language and the associated theory of trust to investigate the interactive structure of trust in a large intraorganizational network. Finally, we extend our analysis to show how social networks can be analyzed in terms of trust neighborhoods, trust corridors, and trust conduits that form a network's preferred conduits for sensitive information. We contribute a way of conceptualizing trust that allows researchers

to detect and measure its presence and degree predictively (by measuring what actors think and think other actors think or know) rather than postdictively (by observing instances of trusting/trustful behavior), and a path for linking trust directly with the ability and tendency of actors in a network to mobilize and coordinate, which are important precursors for cooperation and collaboration.

Trust as a Dependent Variable

Researchers of organizational phenomena have recognized the role of trust in economic life. Some argued that "[t]here is no single variable which so thoroughly influences interpersonal and group behavior as does trust" (Golembiewski and McConkie 1975, p. 131), that "trust is indispensable in social relationships" (Lewis and Weigert 1985, p. 968), and even that trustworthiness is a source of competitive advantage (Barney and Hansen 1994). The idea that trust is critical to the conduct of economic activity has a sociological lineage dating back to the writings of Weber (1992), and in political thought at least to the writing of de Tocqueville (1864, p. 142), who wrote that "the art of association then becomes . . . the mother of action, studied and applied by all." Blau (1964, p. 64) identified trust as "essential for stable social relationships" and Durkheim (1933, p. 28) wrote implicitly of trust when he spoke of the need of a "series of secondary groups" interposed between the state and the individual, that are "near enough to the individuals to attract them strongly in their sphere of action and drag them, in this way, into the general torrent of social life." Fukuyama (1995) conceived "spontaneous sociability" arising from mutual trust as a foundation of economic prosperity, arguing that trust is the lever through which culture influences economic behavior, and "high-trust" and "low-trust" cultures exhibit fundamentally different economic development paths.

Throughout these analyses, however, trust appears implicitly as an explanans of subsequent social behavior rather than an explanandum in need of further elucidation. This situation has motivated several attempts to unpack trust—as a behavioral disposition, as a congruity of mutual expectations, as a special kind of relationship—each informative but also unsatisfactory in one or more ways.

Trust as a Behavioral Disposition

Trust can be seen as the instantiation of cooperative behavior in the prisoner's dilemma game or variants thereof (Lewis and Weigert 1985). Such definitions rely on a view of trust as an individual disposition to behave in a particular way or as a type of individual behavior pattern. It leaves out the interpersonal aspect of trust as a relation between the trusting and the trusted (Barber 1983) as well as the cognitive component of trust, which stresses the mutual expectations of the trusted and the trustful (Gambetta 1988). In the same tradition, trust is also defined in negative terms, as the absence of high-net-benefit opportunistic behavior. Williamson's (1975, 1985) argument, for example, is that economic transactions are structured to eliminate opportunistic behavior on the part of either transactor. Trust appears as a consequence of the absence of opportunistic behavior in the absence of penalties for such behavior.

Behaviorist definitions make no reference to the underlying beliefs of the trusting and the trusted about each other or the situation at hand. Trust must therefore be inferred ex post, from observation of past behavior. Rendering trust a function of past observed behavior alone also rules out investigation of subtler aspects of economic exchange such as forgiveness for certain apparent or real breaches of trust and mutual forbearance in noisy interactional regimes, which limits the precision of the resulting model of trust: When a firm chooses one form of relational contracting over another, we may assume that the rejected contractual form was decided against because of some failure of trust between the actors in question, but we cannot predict the kind of contract that they will enter by reference to the beliefs, expectations, and modes of inference that they use.

Moreover, when behavioral researchers refer to a "probability" of an actor taking a particular action, they refer to the disposition or objective probability of an actor doing so, rather than the subjective probability that colors the perception of his or her observers. This makes trustworthiness—an individual-level variable—the only relevant variable in the analysis of trust, and makes it difficult to analyze situations in which trust is preserved in the face of seemingly guileful behavior or breaks down in spite of apparently principled behavior. Such situations are relevant to the dynamics of trust, however, because the interpersonal interpretation process that leads to the build-up and breakdown of trust is grounded not only in direct behavioral observation, but also in theories, beliefs, and prior interpersonal assumptions that influence what constitutes the relevant observations (Kramer 1999).

Trust as an Interpersonal Entity

The interpersonal dimension of trust is signaled by authors focused on the relationship between trustfulness and trustworthiness (Barber 1983) and by researchers interested in creating comprehensive measures of trust in a relationship (Butler 1991, Mayer et al. 1995). Interpersonal approaches to trust are better at picking out more nuanced constructs, such as benevolence (Mayer et al. 1995), that are specific to

the trusting/trustful relationship. But, on this relational basis alone, we cannot distinguish between insightful benevolence from someone who knows her benevolence can be taken advantage of, but nevertheless chooses to extend it, and credulous benevolence from someone who does not know that she is being taken advantage of, in which case benevolence is more likely to be a case of gullibility.

To unpack trust at this level, we need to get precise about what actors know and how they know what they know, and interactive reasoning offers a tool that we can use to probe inside such differences. The insightfully benevolent actor knows that the other person knows that she can be taken advantage of, whereas the gullibly benevolent actor possesses no such interactive knowledge. For example, Axelrod (1997) recounts the story of a series of prisoner's dilemma games between a Russian male diplomat and an American female diplomat played under "noisy" conditions in which the observation of a defection of one player by another could have been due either to an error on the part of the reporting procedure or to an intentional defection by the other diplomat. The American diplomat came out significantly ahead, and explained that she defected more frequently than she would have—had there been no noise in the game—because she thought the Russian diplomat would attribute most of her defections to reporting errors rather than to intentional actions on the basis of his own cultural stereotypes about the cooperativeness of women. The Russian diplomat confirmed that he had indeed thought of American women as less inclined to defect than American men, and had therefore attributed the defections to errors in the reporting system.

Trust as a Cognitive–Rational Entity: Congruity of Mutual Expectations of Competence

Friedland (1990) posited the desirability of trustworthy behavior on rational grounds alone: In a game in which each player has a choice between cooperative and uncooperative behavior and the perspective of an indefinite number of future interactions, strategies that started out cooperatively and then retaliated responsively and proportionately for uncooperative behavior were likely to bring greater payoffs than strategies that were either uncooperative from the beginning (exemplifying completely untrusting players) or strategies that were cooperative and unresponsive to uncooperative behavior (exemplifying players who trusted blindly). Hill (1990) argued that because transaction costs shaped the form of economic exchange, and because the lack of trust between two transactors increased transaction costs, a reputation for trustworthiness was economically valuable for an actor who did not know ex ante who his possible future business partners might be. Trustworthy behavior builds up the reputation of an individual for trustworthy behavior, increasing the expected value to prospective partners of future interactions with that individual.

Such cognitive–rational congruity views of trust rest on implicit assumptions about actors' beliefs about each other's beliefs. Suppose A trusts B to carry out a particular action, because she believes B to be a rational person who values his reputation and does not want to damage it by shirking on his obligations. Implicitly, A assumes that B indeed cares about his reputation, and that B believes that A sees B's actions as a signal about B's trustworthiness. If B is rational but believes his actions are only imperfectly observable by A, then he might still rationally act in opportunistic fashion and take advantage of A. If A knows this as well, then she will rationally *not* trust B.

In a trustful/trusting relationship, it is not enough that actors trust one another, however; each must also trust her to trust him, trust her to trust him to trust her, and so forth. A's rational trust of B rests on her beliefs about B's beliefs regarding the situation at hand. In turn, B, if he is rational, must consider whether or not his actions will have the desired signaling value. If he believes that A is ill-disposed toward him and will make a negative attribution about his behavior "no matter what he does," then B may not act cooperatively. If A knows this and believes that B believes that A is ill disposed toward him, she will not expect B to act cooperatively. As Chwe (1999) pointed out, such higher-level beliefs are critical to the joint realization of successful cooperative outcomes among rational actors because, given the usual payoff structure of competition–cooperation decisions (highest payoff for matched cooperative actions, lowest payoff for mismatched actions, medium payoff for matched competitive actions), each actor must believe that the other actor knows the payoffs in question for it to be individually rational for the two actors to cooperate.

This interactive logic can be usefully applied to the common view of trust as "the mutual expectation of cooperative behavior" (Burt and Knez 1995). Focus on the word "mutual": j expects k to act cooperatively, and vice versa. What if j does not expect k to expect her to act cooperatively? Then, there will be situations—neatly captured by Prisoner's dilemma game payoff structures—in which it will be logically inconsistent for j to expect k to behave cooperatively. This will induce a belief structure for the pair in which (a) j trusts k but (b) j does not expect k trusts her and therefore (c) finds it irrational to trust k, by the mutual expectation of cooperation definition, which may or may not lead to a reversal of (a). Persevering in believing (a) has costs for j, who must

come to grips with the resulting proposition: "I am irrational," absent some other reason for continuing to trust k while knowing that k does not trust j. To avoid such problems, the expectation of cooperative behavior must be mutual knowledge between j and k: each must know the other knows it, lest we run into problems of logical coherence.

Now, focus on the word "cooperative." Actor j can have cooperative intentions linked causally to behavior that k may deem counter-cooperative: he may intend to help k, for example, but produce behavior that hurts k. Thus, the precise mapping of observable states of the world onto imputed intentions should be mutual knowledge between j and k as well: j and k should impute the same intentionality to the same observable behaviors they see each other produce. Although this is complicated, it is, by itself, not enough: the expectation of cooperative behavior is not restricted to observed behaviors, but also unobserved behaviors. If I trust you (by the standard definition), I know that you will act cooperatively in circumstances that neither of us has experienced before, i.e., "whatever happens."

Finally, focus on the phrase "whatever happens." To infer that you have acted cooperatively in a situation neither of us anticipated, the rules that you use to map a behavior to an intention must be known to me, and the rules that I use to accomplish the same feat must be known to you; they must also be mutual knowledge.

Trust as a Cognitive–Moral Entity: Congruity of Mutual Expectations of Integrity

The moral approach to trust posits norms as guarantors of cooperative behavior because they are ethically justified: "Trust is the expectation by one person, group or firm of ethically justifiable behavior—that is, morally correct decisions and actions based upon ethical principles of analysis—on the part of the other person, group, or firm in a joint endeavor or economic exchange" (Hosmer 1995, p. 399). Granovetter (1985) grounds trust in expectations of conformity of would-be business partners to shared norms and obligations, and cites with approval Arrow's (1974) idea that economic interactions are grounded in a "generalized morality" that is implicitly agreed upon by members of a society. It is rational to act morally and moral to act rationally—the argument goes—and Arrow proceeds to define morality by a commitment to a body of shared norms of behavior. However, there is an implicit epistemic backbone that functions as a safeguard of valid inference here, which is that it is also rational to assume that everyone else is rational and therefore to believe they will also act morally, by virtue of being rational.

To deepen our understanding of trust, we make its epistemic backbone explicit by asking the following questions: What do interacting individuals have to believe about each other's beliefs to engage in the mutually beneficial behavior that is thought to exemplify a "high-trust" relationship? What is the difference between an "attributed" norm and a "shared" norm? An attributed norm is one that A believes that B shares, when B may or may not in fact share it. Thus, A may be wrong about the attributions she makes about B. A shared norm, in contrast, is one that A and B each subscribes to and correctly believes that the other subscribes to as well, that the other believes that she or he subscribes to, and so forth.

Even so, ambiguities remain, and they can only be resolved by further epistemic analysis: It is difficult for any observer of an action to say unambiguously just what norm or maxim to which an action conformed (Moldoveanu and Stevenson 1998). The very same action taken by an actor may be unjustifiable in one moral system (deontological, say) yet remain justifiable in another (utilitarian). Observers' interpretation of each others' actions matter. Whether or not A believes that B was acting in conformity to a particular norm (N) and not some other norm (M) is critical to A's decision, having observed B's behavior, that B is to be trusted. In turn, B's belief that A believes B is acting in accordance with N (rather than M) is critical to B's interpretation of A's retaliatory behavior as a breach of trust in its own right or as a rightful retaliation for A's transgression, by his actions, of norm M. Finally, A's belief that B believed A believed that B was acting according to N rather than according to M will be critical to A's subsequent conciliatory strategy and her reinterpretation of B's behavior accordingly. A detailed consideration of the epistemic states of actors and of the interactive epistemology of their predicaments can usefully inform the way we study rule- and norm-based behavior that is mediated by trust.

Interactive Reasoning as an Integrative Schema for Explanations of Trust

Interactive reasoning forms a basis for unpacking the logic of trust, regardless of whether a moral or rational basis is used as an interpretive schema for trustworthiness, and it also serves as a basis for distinguishing between them. Expecting that trustworthy behavior is in the best interest of a would-be partner or that the would-be partner shares the trust-based ethic of the actor are two plausible grounds for expecting the would-be partner to be trustworthy. Either can work as an explanation of trust. Each will work in different ways, however. As a result, if A trusts B on rational grounds, and B trusts A on moral grounds, the resulting interaction may well produce coordinative disaster and mutual recrimination.

To make progress on disambiguating this explanatory conundrum, we distinguish between "trust in integrity," corresponding roughly to the moral view of trust, and "trust in competence," corresponding roughly to the rational view of trust, and argue that the conjoint condition, trust in integrity and trust in competence, is explanatory of a wide range of trust phenomena. We define trust in competence as a state of epistemic certainty or knowledge of one actor that another would know a proposition if that proposition were true (and relevant to their interaction), and trust in integrity as a state of epistemic certainty or knowledge that an actor would assert that he or she knows to be true. We analyze different combinations of trust in integrity and trust in competence, and examine the differences these differences make. Conceiving of trust in competence and trust in integrity as epistemic states of actors in a network makes it possible for researchers to study trust explicitly by asking actors questions like "what do you believe?" or "what do you think X believes?" rather than implicitly and circularly by asking questions like "whom do you trust?" and to study it predictively (by inferring trusting/trustful behavior from response patterns) rather than postdictively (by inferring them from past behavior).

Because of the precise language in which trust phenomena are represented, however, we must pay special attention to the kinds of propositions that we focus on: after all, one would not expected a trusted *alter* to either know or state *anything* that is true, nor to know everything there is to know about the trustful *ego*. We therefore define a range of applicability of our model of trust by focusing on the propositions that are *relevant* to each of the actors. To do so, we focus on propositions that are important to the comobilization and coordination of two or more actors in a network. We show that what matters to coordination and mobilization in a wide range of contexts is what actors think about what other actors think and about what other actors think they think about a situation. We accordingly posit that trust in integrity and competence should be concomitant to a more coherent set of second-level ("I think you think") and third-level ("I think you think I think") beliefs, because actors who trust one another will be more likely to exhibit coherent interactive belief hierarchies, wherein what *ego* thinks *alter* thinks matches what *alter* thinks, and what *ego* thinks *alter* thinks *ego* thinks matches what *ego* thinks.

Trust in Social Networks

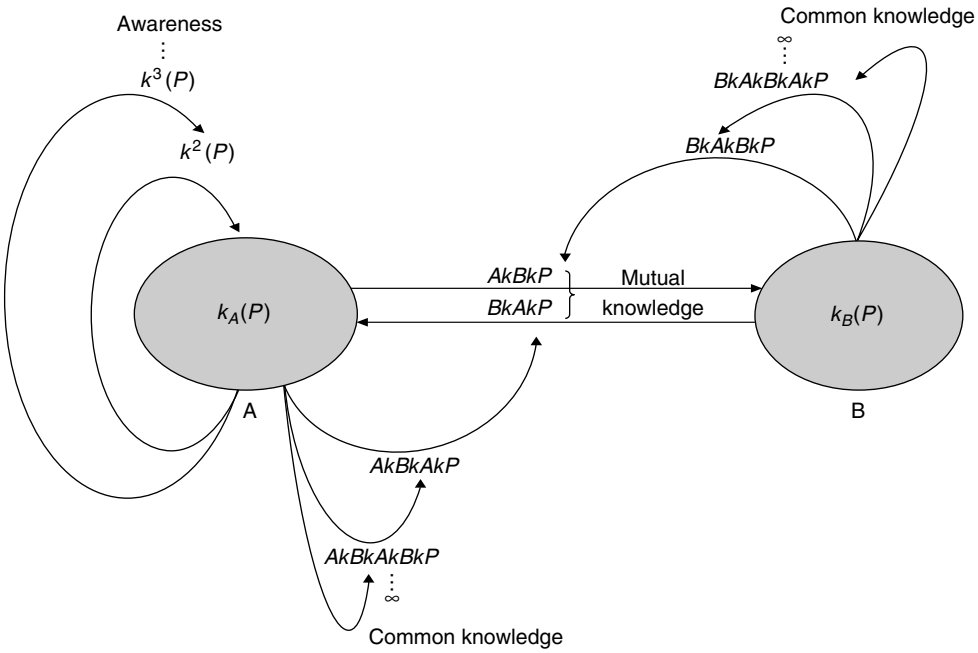
Trust among actors in a network is an important variable affecting social capital (Coleman 1990, Burt 1992, Bourdieu and Wacquant 1992)—or the advantage that accrues from network structure (Bourdieu and Wacquant 1992)—regardless of whether the source of that capital is informational brokerage (exploiting structural holes among densely connected subnetworks) or closure (exploiting densely and redundantly connected subnetworks to enhance the ability of the connected agents to coordinate, cooperate, and collaborate). In a structural hole scenario (Burt 1992), if Alice is the only person that belongs to both subnetwork I and subnetwork II, then her brokerage advantage will be related to the degree to which members of network I (II) trust Alice to convey accurate and reliable information to members of subnetwork II (I). In a closure scenario (Coleman 1990), the fact that Amanda, Bob, and Clarinda all know enough about one another and interact frequently enough for their mutual ties to be considered strong will be far more relevant to their ability to coordinate and cooperate with one another if their mutual acquaintanceship and interactions lead them to trust one another than if they in fact do not. We will build on the insight that "trust moderates social capital" to show how an informational and epistemic view of trust can be used to probe into the nature of the epistemic states of networked agents (both what they know and how they know it) to make predictions about the relative value of brokerage and closure to social capital and to unpack the mechanisms by which differences in brokerage and closure scenarios make a difference to the social capital of network actors.

An Epistemic Description Language for Networked Actors

We specify an epistemic description language for characterizing the individual, collective, and interactive knowledge states of actors within a social network and use it to formalize the notion of an actor's epistemic state. This enables us to represent the knowledge states of actors within a network and demonstrate the importance of actors' individual, collective, and interactive states of knowledge to the achievement of successful coordinative action within the network. Using this description, we define trust in integrity and trust in competence in terms of an underlying set of epistemic states and show how the resulting definition of trust helps us make progress on understanding trust precisely and predictively.

An epistemic state is a relationship of an actor to a piece of knowledge, denoted by a proposition P, or to an experience or intuition E that can be propositionalized by P. Propositionalizability is to be understood here in the counterfactual conditional (or subjunctive) sense: If experience E were to occur, then proposition P would be considered true; otherwise, P would be considered false. Epistemic states of actors are defined as follows (and are summarized in Figure 1).

Figure 1 Individual and Collective Epistemic States



Individual Epistemic States

Knowledge (K)**.** A knows P (AkP) for some actor A and some proposition P. If actor A knows, for instance, that P is "industry-level profits are nondecreasing in industry concentration" then actor A will, ceteris paribus, act as if P is true in those cases in which (a) P is relevant and (b) A sees P as relevant (in which case we will say that P is salient to A). k is a simple binary relation (either AkP or $\sim AkP$) between an actor (or his mind) and a proposition P (or P-propositionalizable experience E) that is characterized by the following necessary conditions: (a) A believes P for reason R, which is valid independently of P being true, and (b) P is true. Note that these are not *sufficient* conditions for knowledge, but merely necessary ones (Gettier 1963).

Confidence. A knows that if P were true, A would know P: $Ak(P \in T \rightarrow AkP)$. Confidence captures A's belief about his epistemic capabilities. It is a binary measure (which can be turned into a continuous measure) that can be used to determine the degree of A's belief in his own epistemic states and answer the question, "Does A believe what she knows?" Note that the converse question, "Does A know what she believes?" is captured by *A's awareness* of her own beliefs, which can be represented as her (second-level) knowledge of her belief or knowledge of a proposition (she knows that he believes or knows it).

Collective Epistemic States

Relevant epistemic states are not limited to those of individual actors: Collective, shared states of knowledge are critical to communication and coordination in networks. We consider a network of actors modeled by a not necessarily fully connected graph G and distinguish among the following network-level epistemic states:

Distribution or Sharedness. A knows P and B knows and C knows $P\ldots$ (AkP & BkP & CkP & $DkP\ldots$). Distribution measures the spread of knowledge of P throughout a network G of actors. It can be measured in absolute terms (the total number of actors that know P) or in relative terms (the proportion of actors in the network G that know P).

Mere distributed knowledge is open to challenges such as imperfect recall (actor k knows P but has temporarily forgotten it) and sub- or unconsciousness (P is not part of k's current working memory), which is why we introduce the following.

Collective Awareness of Level n**.** In this state, A knows^n P&B knows^n P&C knows^n $P\ldots$. Collective awareness measures the degree of level n knowledge about P in the network G and can be operationalized as the proportion of agents that know that they know that they know (up to n levels) some proposition P as a function of n. It can also be operationalized in absolute terms, as the number of actors in G that $know^nP$. Second-level collective awareness (k^2) is a condition that guarantees that the actors in G that know P also will find P salient when P is relevant (because they *know* that they know P). As is the case with knowledge distribution, awareness is not and interactive epistemic property of a network: it simply guarantees that the actors in a network know P and find

P salient when P is relevant. Awareness aids coordinated action in mobilization-type models but still cannot vouchsafe it, because an interactive knowledge structure is still required to induce action. So as not to make the discussion too cumbersome, we will refer to "knowledge" rather than "awareness" henceforth unless otherwise noted.

Interactive Epistemic States

To achieve a characterization of epistemic structures in a network that serves our analysis, we require a representation of the *interactive* epistemic conditions that obtain when we consider *ego's* beliefs about the beliefs of *alters*.

Near Commonality of Level n (NC^n)**.** A knows P, B knows P, A knows B knows P, B knows A knows P, and so forth, to level n: (AkP) & (BkP) & $(AkBkP)$ & $(BkAkP)\ldots$ (abbreviated $(AkBk)^n$ $(P)\&(BkAk)^n(P)$). *Commonality of level* n measures the level of mutual knowledge relative to P of the actors in G. It is a measure of the *coordinative potential* of the network in the sense that many coordination scenarios require not only actor-level knowledge of focal points (Schelling 1978) or coordinative equilibria and a selection criterion among them, but also actor-level knowledge about the knowledge of other actors and about the knowledge that other actors have of the actor's own knowledge (Schelling 1978). Level 2 commonality (mutual knowledge) and level 3 commonality $(AkBkAk(P)\&BkAkBk(P))$ of knowledge about some proposition P are therefore particularly important for studying network-level mobilization and coordination. The value of closure to actors in a densely connected network will be importantly impacted by their ability to accurately and reliably comobilize and coordinate their activities for a common purpose. For this reason, we focus our analysis of trust on propositions that are prima facie important for mobilization and coordination, and are therefore *relevant* to the actors in a network.

Commonality (C)**:** NC^n **with** $n = \infty$**.** A knows P, B knows P, A knows B knows P, B knows A knows P, and so forth, as n increases without bound: (AkP) & (BkP) & $(AkBkP)$ & $(BkAkP)\ldots$. This is the full-common-knowledge state that is usually assumed as a precondition for the justified deployment of the explanatory apparatus of game-theoretic analyses (Brandenburger 1992), and usually without the articulation of a mechanism by which knowledge becomes common or almost common. Common knowledge of strategies, payoffs, and rationality is part of the epistemic conditions for Nash equilibrium (Aumann and Brandenburger 1995, Brandenburger 1992).

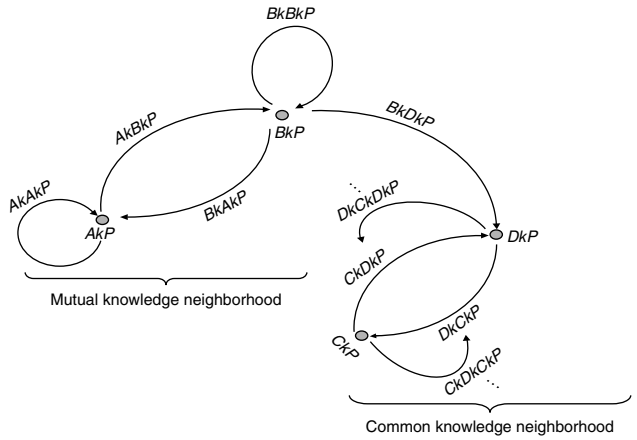
This representation allows us to refer to parts of a network of actors in terms of the degree to which they share epistemically coherent belief hierarchies, as follows.

Mutual Knowledge Neighborhood $(N_{Kn}(G))$**.** This is a subnetwork SG of network G that shares level 2 almost common knowledge of P. Mutual knowledge neighborhoods can be used to describe networks in which there is knowledge about shared knowledge. For example, deliberations of professional and trade associations (American Medical Association, Institute of Electrical and Electronics Engineers conferences) can be understood as turning shared knowledge into mutual knowledge as they unfold: each participant discloses by communicating what he or she knows, which in turn becomes known to the other participants. Mutual knowledge undergirds subnetwork mobilization in situations characterized by "I'll go if you go" scenarios (Chwe 1999, 2000). Suppose A knows "I'll go if and only if (iff) B goes," and B knows "I'll go iff A goes." A will *not* mobilize unless she also knows of B that "he'll go iff she goes," and B will not mobilize unless he knows of A that "she'll go iff he goes." If A does indeed know that B will mobilize if and only if she mobilizes, and B knows that A will mobilize if and only if he mobilizes, then mobilization can take place in the two-actor neighborhood (A, B).

Almost-Common-Knowledge Neighborhood $(N_{SE}(G))$**.** This is a linked subnetwork SG of network G that shares common knowledge of level n of P. "Full-common-knowledge" neighborhoods can be used to model networks in which relevant knowledge P is self-evident. This condition obtains, for instance, in a situation in which P is uttered by one actor in the presence of all other actors, resulting in an epistemic condition in which everyone knows P, everyone knows that everyone knows P, and so forth, ad infinitum. Almost common knowledge of level n is a weaker condition than full common knowledge: if a proposition is common knowledge, then it will also be mutual knowledge (almost common knowledge of level 2 and almost common knowledge of level 3), but not vice versa.

Because commonality of knowledge regarding relevant facts and of the ways in which agents think about these facts is such an important condition for network coordination phenomena, and because it can (at least in theory) be established through face-to-face communication (e-mail trails and exchanges can at best establish almost common knowledge of level n, as Rubinstein 1986 illustrates), understanding it as an epistemic condition that enables coordination highlights the importance of such otherwise curious phenomena as face-to-face meetings in the age of ubiquitous untethered broadband communications. However, almost common knowledge is often a sufficient condition for coordination among several actors. If Aretha and Bertha are trying to meet one another on a small island that has a single large hill, but had neglected to establish a meeting place

Figure 2 Interactive Epistemic States



beforehand, then a sufficient condition for them to solve their coordination problem is that (a) Aretha knows Bertha knows Aretha knows Bertha will head for the hill, and (b) Bertha knows Aretha knows Bertha knows Aretha will head for the hill (i.e., $n = 3$). Thus, second- and third-level almost common knowledge are important conditions for comobilization and coordination in networks, and, accordingly, important preconditions for turning network closure into social capital whose value lies in enabling comobilization and coordination.

Figure 2 graphically illustrates the interactive epistemic states that we have introduced.

Trust as an Interactive Epistemic State

Having established the importance of higher order epistemic states to coordination and mobilization among networked actors, we now turn to an epistemically precise characterization of trust. Our definition makes use of modal epistemic structures such as the one used above to define confidence. As such, trust can be understood as a form of confidence. We decompose trust into two separate entities, trust in integrity (actors' propensity to say that which is true) and trust in competence (actors' propensity to know the truth).

Trust in Integrity (Strong Form) (T_I)**.** A knows that iff B knew P, B would assert P: $Ak(BkP \leftrightarrow BaP)$, where BaP denotes "B acts on P" or "B asserts P," and P denotes a proposition that is relevant to the interaction between A and B. If A trusts B's integrity, then A knows that "iff B knows the truth, then B will assert the truth." In this form, (a) knowledge by B of P is sufficient to guarantee to A that B would say P, and (b) if P were false, then B would not say it, i.e., B speaks the truth and nothing but the truth.

This definition of trust generates the following interactive belief hierarchy between A and B: Suppose P is the proposition "B does not trust A." Assume the proposition is known to B, unknown to A, and (evidently) relevant to the interaction between A and B. If it is true that A trusts B, then A knows that iff B knew that P is true, then B would assert it, and therefore that B would inform A of his mistrust. Because of the biconditional (iff) nature of the relationship between knowledge and expression, A can infer from the fact that B does not say "I don't trust you" the fact that B actually trusts A. The argument is symmetric for A. Thus, trust in integrity as we have defined it generates an interactive belief hierarchy between the parties to the trusting relationship.

Our definition of trust in integrity (T_I) sidesteps two complications. The first is that B must be aware that he knows P for him to say P given that he knows it. The second is that the biconditional $\leftrightarrow$ relating B's knowing P to B's asserting P is plausibly relaxed to the simple conditional, such that B will assert what he knows but will not necessarily know all that he asserts. These complications can be repaired, respectively, by the following relaxations of T_I.

Trust in Integrity (Weak Form 1) (T_{Iwf1})**.** A knows that iff B were (level 2) aware of P, then B would assert P: $Ak(Bk^2P \leftrightarrow BaP)$. This weak form of trust in integrity requires that B's awareness of P be independently ascertained by A: knowledge by A of knowledge by B of P is not sufficient to warrant A's knowledge that B would assert P. Thus, if B has reason to know P and A knows it and weak-form trusts B, then B's failure to bring up P is forgiveable in the sense that A can continue to trust B in the weak sense, provided A considers it possible that B temporarily overlooked or forgot P. This weaker form of trust incorporates imperfect recall, which has an important explanatory function in some interactive analyses of games (Aumann 2001).

Trust in Integrity (Weak Form 2) (T_{Iwf2})**.** A knows that if B knew P, B would say P or act as if P were true: $Ak(BkP \rightarrow BaP)$, where BaP denotes "B acts on P" or "B asserts P." If A trusts B's integrity in this weak sense, then A knows that "if B knows the truth, then B will assert the truth," but not necessarily that all B would assert is the truth. B, in other words, is "trusted" by A to assert the truth, but not to assert "nothing but the truth."

The second part of our definition of trust refers to trust in competence. A may trust B's intentions, but not necessarily B's ability to deliver, perform, or make good on those intentions. In the way we have defined trust, B is only expected by A to say what he knows to be true, but not to know that which is true. Thus, we extend our definition of trust as follows.

Trust in Competence (T_C)**.** A knows that iff P were true, B would know P: $Ak(P \in T \leftrightarrow BkP)$. If A trusts in the competence of B, then A will trust that B is a

faithful register of true propositions and not a register of false propositions, i.e., that B will register the whole truth.

Now we are in a position to characterize trust as the combination of trust in integrity and trust in competence. In the strong form, trust combines the strong form of trust in integrity (which we shall use henceforth unless we signal otherwise) with trust in competence, as follows.

Trust Tout Court (T)**.** A trusts in both the competence and integrity of B: $ATB \rightarrow (AT_IB \& AT_CB)$. If ATB, then A knows that B will assert the truth, the whole truth, and nothing but the truth.

Weaker forms of trust can be created by combining the trust-in-competence condition with the weaker forms of the trust-in-integrity condition, permitting us to create "trust hierarchies" and, rather than declaring trust to exists or not within a relationship, to become far more precise regarding the kind of trust that exists between two or more actors. Moreover, for any or all of a collection of actors A, B, and C, the relation T has the following properties.

Binariness B. $\forall A, B$: ATB or $\sim ATB$. For any A and B, A either trusts or does not trust B.

Proof. Focus first on T_I. Suppose that AT_IB and $\sim AT_IB$, i.e., $Ak(BkP \leftrightarrow BaP) \& Ak(BkP \leftrightarrow \sim BaP)$. But this contradicts the binariness property of the k relation (a). The same argument applies for T_C, and therefore for the conjunction $T = T_I \& T_C$.

Transitivity T. $\forall A, B, C$: ATB & $BTC \rightarrow ATC$. For any A, B, and C, if A trusts B and B trusts C, then A trusts C.

Proof. If AT_IB, then $Ak(BkP \leftrightarrow BaP)$. If BT_IC, then $Bk(CkP \leftrightarrow CaP)$. Let $P =$ "$CkR'' \leftrightarrow CaR$" for some R. Clearly, BkP and, ex hypothesis, BaP. Because A knows that B will only say what B knows (to be true) and knows to be true only those sentences that are in fact true, AkP, and therefore AT_IC. Now, let $Q = R \in T \leftrightarrow CkR$. Clearly, BkQ, and given AT_IB, $Ak(Q \in T)$. Therefore, AT_CC. Therefore, AT_CC and AT_IC, i.e., ATC.

Does the fact that A trusts B entail the fact that B trusts A? The answer is no, and the proof is by counterexample. Suppose A trusts B to assert "the truth, the whole truth and nothing but the truth," but B does not trust A to assert the truth, the whole truth and nothing but the truth. Let P represent the proposition "B does not trust A to assert the truth, the whole truth and nothing but the truth." *Ex hypothesis*, P is true. Since A trusts B, B will assert P and A will know P, and therefore will know that B does not trust her. So, if A trusts B and B does not trust A then our definition requires A to know that B does not trust A. However, in order for A to not trust B as a result of this knowledge, it is necessary to make A's trust in B conditional upon A's knowledge of B's trust in A, which it does not necessarily have to be: A may trust B on account of the "kind of person A believes B is" rather than on account of a set of expectations about B's behavior given a set of incentives and a purely self-maximizing disposition. Indeed, if we introduce a self-maximizing interpretation for the foundation of trust (a mutual expectation of cooperative behavior in situations with payoffs captured by the Prisoner's Dilemma game played between self-interested individuals; e.g., Axelrod 1997) then it will be the case that A cannot trust B coherently while knowing that B does not trust A. The proof is simple and follows directly from the application of the definition of trust to the epistemic state space of the *PD* game.

Trust, Coordination, Closure, and the Coherence of Interactive Beliefs

Our definition of trust is obviously too broad if it is applied to all propositions that are possibly true: A cannot reasonably know or expect that B will know and assert any true sentence. The range of admissible propositions, then, should be restricted (just as in our definition of knowledge) to propositions that are *relevant* to the interaction between A and B. Coherence of higher-order beliefs among networked actors is a condition for successful coordination and mobilization, and therefore propositions that are the focal points of coordination and mobilization scenarios are particularly interesting to our model of trust: these are the propositions that are most likely to be "relevant" to both A and B.

Interactive belief hierarchies are interesting even if we do not go "all the way to common knowledge" and focus on level 2 and level 3 beliefs (Ayres and Nalebuff 1997). In spite of the fact that full common knowledge cannot be achieved through non-face-to-face communication (as in Rubinstein's (1986) e-mail game), most human creatures *infer* common knowledge from second-order knowledge when it comes to confirming a meeting by e-mail: they do not "confirm confirmations" or "confirm confirmations of the confirmations," but they nonetheless wait for confirmation before assuming that a meeting will take place at the proposed time and place. Nov and Rafaeli (2009) have shown that, in organizational contexts, individuals attach a premium to mutual knowledge (although the authors mistakenly call it "common knowledge") relative to shared knowledge. This highlights the value that individuals place on the interactive component of interactive belief hierarchies in coordination-intensive activities. Culture, whether social or organizational, may thus, as Kreps (1990) and Chwe (2000) have explained, represent an interactively coherent body of shared beliefs: they may either

be common knowledge or almost common knowledge among the individual members thereof.

It seems reasonable, therefore, to focus on the coherence of interactive beliefs—Does what I believe match what you believe I believe? Does what I believe you believe I believe match what I believe?—as indicative of the mobilize-ability and coordinate-ability of groups of actors, and therefore of a significant component of the value of closure in human networks. If almost common knowledge serves the crucial purpose of enabling coordination on interdependent activities, and if we assume that individuals are more likely to engage in coordinated action with other individuals they trust than with those they do not, then we should be able to observe a link between trust and coherence of interactive beliefs.

If trust engenders hierarchies of coherent level 2 and 3 beliefs, then we should observe a high correlation between coherent belief hierarchies and trust. Moreover, the correlation may be stronger than that between coherent belief hierarchies and, for instance, the relative centrality of linked actors (which often proxies for their informedness, but not about their ability to successfully coordinate and comobilize with other actors), or the relative number of shared ties among actors' alters within the network as measured by their constraint. It is commonly *assumed* that network centrality brings with it an enhanced ability to mobilize or coordinate the network. Whatever the source of this enhanced ability, the arguments we have advanced indicate that it must possess an epistemic component, because mobilization and coordination are network processes that have to rest on second- and third-level almost-common-knowledge conditions: Mobilizing actors must have accurate information about what other actors know, whereas coordinating actors must have accurate information about both what other actors know and about what other actors know they know. Density of relationships too is widely held to influence the knowledge distribution and thus mobilize-ability and coordinate-ability of a network. Chwe (1999, 2000), for example, takes cliques (fully connected subnetworks) to be synonymous to pockets of local common knowledge. But they need not be; one may communicate with several others who are all communicating to each other (i.e., be part of a clique), yet the mobilizability of each actor may not come to be common knowledge among clique members. Nor, as we have shown, is full common knowledge necessary in many mobilization scenarios; rather, mutual knowledge of comobilization is sufficient.

Data and Method

Our epistemic definition of trust allows us to investigate trust relationships empirically and noncircularly: empirically by uncovering what actors believe about each other's propensity and know and speak the truth to each other, and noncircularly by using the definitions of trust in competence and trust in integrity to probe into the epistemic and logical structures of trust without ever asking "whom do you trust?"

To examine these insights empirically, we collected network and epistemic data from senior managers of a large multidivisional, multiregional Canadian telecommunications firm. The data were collected in two phases. The first entailed collection of network and biographical data through an online survey. Respondents were invited to complete the survey through a personalized e-mail that contained a link to the survey. The survey asked respondents to identify their significant work relationships, without restriction on the number of colleagues they could mention. To ensure accuracy (e.g., consistency in spelling; to distinguish people with similar or identical names), when respondents entered the name of a colleague in the survey, the name was checked against a database containing the names and titles of all the firm's employees, and a list of suggested names and titles was presented to the respondent from which she or he could select the desired individual. For each work relationship identified, the respondent was asked to rate the strength of the relationship from 1 (very weak) to 5 (very strong). In addition to their significant relationships, each respondent indicated their geographic location, corporate division affiliation, and gender. Of the 633 individuals invited to participate in the online survey, 593 (93.7%) completed it. The high response rate was obtained as the result of strong support from top management and individualized e-mail reminders from the researchers to complete the surveys.

We used these network data to measure each respondent's betweenness centrality and constraint. Betweenness centrality measures the extent to which an actor lies on the shortest paths between other actors in the network (Freeman 1977). Betweenness for respondent i is $\Sigma_{jk}\sigma_{jk}(i)/\sigma_{jk}$, where σ_{jk} is the number of shortest paths from j to k, and $\sigma_{jk}(i)$ is the number of shortest paths from j to k that pass through respondent i. Constraint is a measure of the cohesiveness of relationships surrounding an actor (Burt 1992). A respondent's ego network is constraining to the extent that it is directly or indirectly concentrated in a single contact; more constraint implies fewer bridging ties. Constraint for respondent i is $\Sigma_j(p_{ij}+\Sigma_q p_{iq}p_{jq})^2$, $q \neq i, j$, p_{ij} is the strength of respondent i's relationship with alter j, and p_{iq} and p_{jq} are the strengths of alter q's relationships with i and j, respectively. When $p_{iq}p_{jq}$ is large, a strong third-party tie connects respondent i to alter j indirectly. Summing over q provides an assessment of the overall strength of third-party

ties surrounding respondent i. Constraint varies from a minimum of p_{ij}^2 when alter j is disconnected from all of respondent i's other alters, to a maximum of 1, when j is i's only alter.

The second phase of data collection took place during a series of voluntary workshops to which all 633 individuals were invited to receive feedback on their networks. Each workshop attendee received a personalized survey to complete. To assess first-order (level 1) beliefs, the survey asked respondents to indicate whether they believed their firm's success depended most importantly on "innovation," "focus," or "marketing."[1] Respondents were also asked to answer the same question for alters (to a maximum of 10) with whom they rated their relationship strength either 4 or 5 (on the five-point scale) on the online survey. For each of these alters, respondents were also asked whether they believed (1) the alter would know what they believed the firm's success depended on (second-order or level 2 beliefs), and (2) the alter believed they would know how the alter responded (third-order or level 3 beliefs). Propositions about the drivers of the success of the firm were selected for our queries because they are relevant to both the successful coordination and comobilization of the respondents; they can be thought of as focal points in a game in which managers try to coordinate their actions around a set of accepted goals. To the extent that these goals are common knowledge or almost common knowledge, they form an important part of the "culture" of the organization (Kreps 1990).

Respondents rated the strength of their beliefs from 1 (definitely not) to 7 (definitely). Finally, to assess trust in competence and integrity, the respondents were asked to rate, on a seven-point scale, whether they could count on each of the alters "to be 'in the know' and to communicate what she or he knows."[2] A total of 302 survey responses were received from workshop attendees, of which 296 were useable for the analysis. Combining the network and epistemic survey data yielded 608 significant work relationships among the survey respondents, of which 113 were reciprocated at the required cutoff level for both respondents.[3]

Results

Table 1 presents descriptive statistics and bivariate correlations among the study variables for the sample of 608 significant work relationships. Of particular interest are correlations among respondent i's trust in the competence and integrity of alter j, level 2 and 3 beliefs regarding alter j, knowledge of alter j's level 1 belief about the basis of the firm's success (the variable *ij matched*, coded 1 if correct, and 0 otherwise), and measures of respondent i's betweenness centrality and constraint. The correlations are strong and positive among the strength of trust and level 2 and 3 beliefs, as well as with knowledge of alters' level 1 beliefs regarding the basis of firm success. Betweenness centrality and constraint are weakly correlated with these variables, however, and with one exception not significantly different from zero.

Table 2 reports ordinary least squares (OLS) regression estimates for a hierarchically nested set of multivariate models of trust. These models estimate each variable's independent association with trust, which simple correlations do not. They also permit inclusion of additional control variables that may affect respondents' trust in their alters, including the respondent's and alter's gender and tenure with the firm, whether the respondent and alter are the same gender, and whether the respondent and alter work in the same location and/or company division. Because our theoretical variables may be correlated with other general types of information that the survey respondents may possess about their alters, and because our theory relies on specific epistemic states, not general information about alters, it is important to control for these alter characteristics, which are plausibly correlated with the specific level 1 and 2 beliefs we studied.

Given the correlations among the variables, we report a set of hierarchically nested models to check whether multicollinearity was imposing a conservative bias on our estimates by inflating coefficient standard errors. The absence of such inflation is reinforced by variance inflation factor (VIF) statistics for each model, which reach a maximum of 1.34, well below the threshold of 10 (Belsley and Welch 1980). The regression estimates confirm the correlational analysis. Coefficients for both level 2 and level 3 beliefs are significant and positive, whereas the network variables are not. Notably, common work location and alter tenure are also positively related to respondents' trust in the competence and integrity of alters.

Table 3 reports multivariate Logit regression models estimating whether or not respondents' were correct about alter j's level 1 beliefs about the basis of their firm's success (*ij matched*). These models estimate the independent associations of respondents' trust of alter j, level 2 and 3 beliefs about alter j, and network positions with respondents' correct identification of alter j's level 1 beliefs, controlling for respondent and respondent and alter demographic characteristics. Again, neither nested model estimates nor VIF statistics reported for each model indicate

[1] These factors were selected by the firms' senior management.

[2] Ideally, this final item would have been separated into two questions, one focused on competence and one on integrity. They were combined at the urging of the firm's senior management to shorten the survey.

[3] Because we wanted to focus on significant ties, we did not include highly asymmetric reciprocal ties (i.e., where one respondent rated the tie to be of low significance and the other of high significance).

Table 1 Bivariate Corrections of All Relationships

	Descriptive stastistics and bivariate correlations																	
	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 *i trust level*	5.82	1.23	1	7	1.00													
2 *i level 2 belief strength*	5.30	1.26	1	7	0.40	1.00												
3 *i level 3 belief strength*	5.35	1.29	1	7	0.40	0.66	1.00											
4 *ij matched*	0.59	0.49	0	1	0.17	0.25	0.17	1.00										
5 *i betweenness*	0.0017	0.0019	0.0000	0.0127	−0.01	0.08	−0.05	0.06	1.00									
6 *j betweenness*	0.0016	0.0018	0.0000	0.0127	0.01	0.03	0.00	−0.01	−0.02	1.00								
7 *i constraint*	0.093	0.056	0.017	0.439	−0.01	−0.04	0.01	0.04	−0.53	0.05	1.00							
8 *j constraint*	0.100	0.059	0.017	0.656	−0.03	−0.01	0.00	0.06	0.01	−0.53	0.01	1.00						
9 *ij same location*	0.61	0.49	0	1	0.12	0.07	0.13	0.00	−0.10	0.07	0.04	0.06	1.00					
10 *ij same division*	0.61	0.49	0	1	0.05	0.02	0.12	0.08	−0.14	−0.05	0.10	0.08	0.04	1.00				
11 *ij same gender*	0.69	0.46	0	1	0.06	0.11	0.12	−0.04	−0.14	−0.03	0.12	−0.01	0.11	0.01	1.00			
12 *i male*	0.68	0.47	0	1	0.08	0.12	0.06	−0.03	−0.16	0.01	0.10	−0.03	0.14	−0.10	0.38	1.00		
13 *i tenure*	12.00	9.67	1	35	0.02	0.09	0.08	0.01	−0.06	0.01	0.07	0.06	−0.02	−0.11	0.03	−0.04	1.00	
14 *j male*	0.73	0.44	0	1	0.02	0.10	0.11	0.02	−0.08	−0.04	0.06	0.09	0.11	0.00	0.23	0.26	−0.02	1.00
15 *j tenure*	12.13	9.70	1	35	0.07	0.04	0.03	0.02	−0.10	0.01	0.10	−0.05	0.01	0.03	0.05	−0.02	0.29	0.01

Note. $N = 608$; correction, 0.08; significant at $p < 0.05$.

multicollinearity concerns. In the models, coefficients for both trust and level 2 beliefs are significant and positive, whereas coefficients for level 3 beliefs are not. Notably, in the full model, consistent with the idea that network centrality and density of relationships influence the knowledge distribution (e.g., Burt 1992; Coleman 1988, 1990), respondents' whose network positions are more central and constrained respondents are more likely to be knowledgeable about their alters' level 1 beliefs. Common work division is also positively related to knowledge of alters' beliefs, which is sensible in light of the nature of beliefs assessed.

The foregoing analysis is based on all significant work relationships, regardless of whether or not they were reciprocated. We now turn our attention to relationships that 113 reciprocated relationships, which permits a more fine-grained dyadic analysis of the effects of trust and belief strength.

Table 4 presents descriptive statistics and bivariate correlations among the study variables for the sample of reciprocated relationships. Of particular interest are respondents' trust in the competence and integrity of and level 2 and 3 beliefs about each other, as well as their network characteristics and demographic characteristics. As before, the correlations are strong and positive for respondents' trust in and level 2 and 3 beliefs about alters. The correlation between each respondent's trust in the other is also strong and positive. The correlations between respondents'

Table 2 OLS Regression Models of Trust

	OLS estimates									
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
i trust level										
ij same location	0.25	(0.10)**	0.26	(0.10)**	0.26	(0.10)**	0.26	(0.10)**	0.18	(0.09)*
ij same division	0.16	(0.10)+	0.17	(0.10)*	0.17	(0.10)*	0.18	(0.10)*	0.05	(0.09)
ij same gender	0.03	(0.12)	0.03	(0.12)	0.03	(0.12)	0.03	(0.12)	−0.02	(0.10)
i male	0.17	(0.12)+	0.19	(0.12)+	0.17	(0.12)+	0.18	(0.12)+	0.10	(0.11)
i tenure	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.00)
j male	−0.03	(0.12)	−0.02	(0.12)	−0.02	(0.12)	−0.01	(0.12)	−0.12	(0.10)
j tenure	0.01	(0.01)+	0.01	(0.01)+	0.01	(0.01)+	0.01	(0.01)+	0.01	(0.00)*
i betweenness			23.88	(27.31)			22.54	(31.65)	−1.70	(27.88)
j betweenness			2.65	(27.91)			−10.80	(33.06)	−12.23	(28.88)
i constraint					−0.51	(0.91)	−0.11	(1.05)	−0.25	(0.92)
j constraint					0.59	(0.86)	−0.79	(1.01)	−0.73	(0.88)
i level 2 belief strength									0.23	(0.05)***
i level 3 belief strength									0.21	(0.05)***
Constant	5.38	(0.16)***	5.30	(0.19)***	5.46	(0.19)***	5.40	(0.25)***	3.46	(0.28)***
Model VIF	1.12		1.11		1.11		1.24		1.34	
F-statistic	2.10*		1.71+		1.71+		1.45		11.96***	
Adj. R^2	0.01		0.01		0.01		0.01		0.19	

Note. $N = 608$.

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 3 Logit Regression Models of Knowledge of Alters' Level 1 Beliefs

	Logit estimates													
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
ij matched														
ij same location	0.01	(0.17)	0.03	(0.17)	−0.01	(0.17)	0.03	(0.17)	−0.06	(0.18)	−0.09	(0.18)	−0.07	(0.18)
ij same division	0.34	(0.17)*	0.39	(0.17)**	0.31	(0.17)*	0.37	(0.18)*	0.33	(0.18)*	0.27	(0.18)$^{+}$	0.30	(0.18)*
ij same gender	−0.22	(0.20)	−0.20	(0.20)	−0.23	(0.20)	−0.22	(0.20)	−0.23	(0.20)	−0.27	(0.21)$^{+}$	−0.30	(0.21)$^{+}$
i male	−0.04	(0.20)	0.01	(0.20)	−0.05	(0.20)	0.01	(0.20)	−0.05	(0.21)	−0.04	(0.21)	−0.14	(0.21)
i tenure	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)
j male	0.16	(0.20)	0.17	(0.20)	0.13	(0.20)	0.14	(0.20)	0.16	(0.20)	0.10	(0.21)	0.08	(0.21)
j tenure	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)
i betweenness			91.12	(48.83)*			153.52	(59.07)**	150.60	(59.90)**	153.69	(60.26)**	121.56	(60.00)*
j betweenness			−8.12	(46.69)			24.78	(56.69)	27.33	(57.26)	24.11	(57.24)	15.97	(57.71)
i constraint					3.34	(1.58)*	3.99	(1.97)*	4.19	(2.00)*	4.22	(2.00)*	4.24	(2.02)*
j constraint					1.90	(1.54)	2.25	(1.89)	2.47	(1.88)$^{+}$	2.40	(1.85)	2.36	(1.86)
i trust level									0.32	(0.07)***	0.22	(0.08)**	0.16	(0.08)*
i level 2 belief strength													0.38	(0.10)***
i level 3 belief strength											0.21	(0.07)**	0.00	(0.09)
Constant	0.19	(0.27)	−0.08	(0.32)	−0.05	(0.31)	−0.76	(0.44)*	−2.52	(0.60)***	−2.93	(0.63)***	−3.28	(0.65)***
VIF	1.12		1.11		1.11		1.24		1.23		1.25		1.35	
Log-likelihood	−409.60		−407.74		−408.43		−404.61		−394.06***		−389.21***		−381.44***	

Note. $N = 608$.

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

level 2 and 3 beliefs and their trust in each other are also positive, although more weakly, and significant only in the case of level 3 beliefs. Three of four correlations among respondents' level 2 and 3 beliefs are positive and significant as well.

Among the network measures, each respondent's constraint is positively correlated with her or his own level 2 belief strength, consistent with the idea that the density of relationships promotes interpersonal cohesion and formation of norms (e.g., Coleman 1988, 1990). More strikingly, however, respondents' strength of trust in and level 2 and 3 beliefs about alter *j* are negatively and significantly correlated with alter *j*'s betweenness centrality. These negative correlations are consistent with the idea that network centrality fosters a competitive orientation among actors as they attempt take advantage of opportunities for information brokerage and control to increase their autonomy and others' dependence on them (Burt 1992, Moldoveanu et al. 2003).

Table 5 presents a multivariate analysis of trust in the reciprocated relationships. Because errors are likely to be correlated across equations estimating the trust of respondent *i* on alter *j*, and alter *j* on respondent *i* using the same data, we estimate these equations using a seemingly unrelated regression (SUR) model, which allows correlated errors between equations (Greene 2000). The multivariate analysis again confirms the correlation analysis. Consistent with the estimates in Table 3, estimates based on reciprocated relationships indicate that the respondents' level 2 beliefs about alters are positively associated with their trust in them, whereas level 3 beliefs are not. Estimates also show respondents' independently reported trust in each other is positively related. Respondents' network constraint is also positively correlated with their level 2 belief strength. The estimates also indicate that respondents' trust in alters is negatively related to alter betweenness and positively related to alter constraint. Again, these findings are consistent with the idea that the density of relationships promotes interpersonal cohesion and norms of cooperation (e.g., Coleman 1988, 1990), whereas central network positions create opportunities and incentives for information brokerage and control (e.g., Burt 1992, Moldoveanu et al. 2003). Among the control variables it is notable, and somewhat ironic, that more senior respondents expressed more trust in alters while simultaneously being less trusted.

Taken together, the results of the foregoing analysis corroborate the idea that trust is associated with deeper hierarchies of coherent level 2 and 3 beliefs. The analysis reveals not only a strong covariation among respondents' level 2 and 3 beliefs about alters and their trust in the competence and integrity of alters, but also with their knowledge of alters' level 1 beliefs. In contrast, although respondents' betweeness centrality and constraint were positively associated with their knowledge of alter *j*'s level 1 beliefs, they were either unrelated or negatively related with their trust in alters. Thus, actor centrality is not necessarily indicative of relative informedness (especially when higher-level knowledge is included in what counts as information) and the enhanced ability to mobilize or coordinate the network that accompanies it, nor is actor constraint necessarily indicative of common

Table 4 Bivariate Correlations for Reciprocal Relationships

					Descriptive stastistics and bivariate correlations														
	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 *ij trust level*	6.10	1.14	1	7	1.00															
2 *ij level 2 belief strength*	5.44	1.30	1	7	0.39	1.00														
3 *ij level 3 belief strength*	5.58	1.07	3	7	0.28	0.43	1.00													
4 *ji trust level*	6.12	1.15	1	7	0.30	0.13	0.17	1.00												
5 *ji level 2 belief strength*	5.50	1.28	1	7	0.13	−0.07	0.18	0.37	1.00											
6 *ji level 3 belief strength*	5.62	1.06	3	7	0.15	0.17	0.17	0.26	0.44	1.00										
7 *i betweenness*	0.0015	0.0017	0.0000	0.0102	−0.07	0.02	−0.12	−0.13	−0.27	−0.32	1.00									
8 *j betweenness*	0.0015	0.0017	0.0000	0.0102	−0.17	−0.29	−0.38	−0.08	0.11	−0.09	0.12	1.00								
9 *i constraint*	0.102	0.056	0.017	0.439	−0.02	−0.09	−0.11	−0.03	0.18	0.10	−0.54	0.09	1.00							
10 *j constraint*	0.100	0.055	0.017	0.439	−0.03	0.20	0.13	−0.03	−0.13	−0.12	0.07	−0.55	0.03	1.00						
11 *ij same location*	0.62	0.49	0	1	0.29	0.20	0.31	0.29	0.20	0.34	−0.16	−0.10	−0.01	−0.04	1.00					
12 *ij same division*	0.64	0.48	0	1	−0.03	0.09	0.07	−0.04	0.09	0.09	−0.14	−0.08	0.20	0.17	0.05	1.00				
13 *ij same gender*	0.76	0.43	0	1	−0.06	0.02	0.19	−0.07	0.01	0.21	−0.37	−0.30	0.04	0.02	0.12	−0.03	1.00			
14 *i male*	0.73	0.45	0	1	0.05	0.18	0.11	−0.02	−0.01	0.02	−0.18	−0.21	0.07	−0.03	0.13	−0.09	0.31	1.00		
15 *i tenure*	12.04	9.85	1	35	−0.03	0.11	0.01	0.18	0.07	0.10	−0.18	−0.20	0.11	0.18	0.09	−0.08	0.18	−0.02	1.00	
16 *j male*	0.73	0.44	0	1	−0.02	0.00	0.07	0.06	0.12	0.09	−0.15	−0.17	−0.03	0.04	0.11	−0.12	0.27	0.39	0.06	1.00
17 *j tenure*	12.16	9.85	1	35	0.19	0.06	0.10	−0.04	0.10	0.01	−0.18	−0.16	0.14	0.13	0.08	−0.09	0.17	0.06	0.38	−0.03

Notes. $N = 113$. Correlations greater than 0.17 are significant at $p < 0.05$.

knowledge. Moreover, when high constraint in networks results from dense connections among actors (rather than connections through a single individual, as in a hierarchy), constraint ought to be positively related to trust (Burt 1992). Although constraint derives primarily from density in our empirical setting, this isn't generally the case in our findings. The model of epistemic states thus does a better job of explaining trust—and predicting it ex ante—than an intuitive and widely accepted structural explanation.

Extensions: Trust and the Epistemic Dynamics of Social Networks

Trust and security are vital to information diffusion, verification, and authentication in social networks and to a precise explanation of the network position advantage that accrues to certain actors. The effective brokerage of information across a structural hole is dependent upon the trustworthiness of the broker to both the transmitter and the recipient of that information. If trust safeguards coordination and mobilization, then it is reasonable to assume that it will also safeguard coordination-intensive tasks such as truthful communication and, therefore, the spread of useful and veridical information within a network. Communicating critical information is a coordination-intensive task because it often takes place on the background of shared assumptions and orientations, whose commonality is important to the accurate receipt of the information in question: "[the CEO] told me we are moving away from vertical markets" may be highly useful information to someone who knows the context of the conversation in which the information emerged and quite useless to someone who does not. In this section, we show how the description language we have introduced can be used to model the flow and authentication of information in a social network and pave the way to empirical analyses of informational dynamics in networks. The epistemic description language we developed allows us to state how and why trust and security matter by making it possible to state sufficient conditions for the knowledge that trusting, trusted, secure, and authenticated networked actors fulfill. In particular, we define the following.

We define *trust neighborhood* $N_{Ti}(G), N_{Tb}(G)$ as a fully linked subnetwork of G (clique) that shares trust in mutual competence and integrity (weak form). A trust neighborhood is a good model for a network of "close ties," wherein actors can rely on one another for truthful and truth-like knowledge sharing, conditional upon awareness: if an actor is aware of P (knows it and knows she knows it), then she will share P with others. Thus, in a trust neighborhood,

Table 5 SUR Estimates of Trust in Reciprocal Relationships

	SUR estimates											
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
ij trust level												
ij same location	0.70	(0.21)***	0.67	(0.20)***	0.69	(0.21)***	0.64	(0.20)**	0.35	(0.20)*	0.30	(0.20)+
ij same division	−0.10	(0.21)	−0.19	(0.21)	−0.07	(0.22)	−0.12	(0.21)	−0.05	(0.21)	−0.11	(0.20)
ij same gender	−0.32	(0.25)	−0.49	(0.26)*	−0.32	(0.25)	−0.54	(0.27)*	−0.28	(0.26)	−0.21	(0.25)
i male	0.10	(0.25)	0.03	(0.25)	0.11	(0.25)	−0.01	(0.25)	0.06	(0.24)	−0.14	(0.23)
i tenure	0.01	(0.01)	0.02	(0.01)+	0.01	(0.01)	0.02	(0.01)+	0.03	(0.01)**	0.03	(0.01)**
j male	−0.06	(0.25)	−0.11	(0.25)	−0.06	(0.25)	−0.09	(0.24)	−0.14	(0.24)	−0.03	(0.23)
j tenure	−0.03	(0.01)**	−0.02	(0.01)*	−0.03	(0.01)**	−0.02	(0.01)**	−0.03	(0.01)**	−0.03	(0.01)**
i betweenness			−49.07	(64.01)			−40.28	(77.25)	21.72	(75.78)	−1.75	(72.18)
j betweenness			−135.44	(61.67)*			−198.63	(75.66)**	−161.16	(73.66)*	−120.45	(73.39)*
i constraint					0.59	(1.85)	0.35	(2.20)	0.60	(2.14)	1.13	(2.01)
j constraint					0.37	(1.88)	3.49	(2.25)+	2.77	(2.19)	3.58	(2.07)*
ji trust level									0.46	(0.09)***	0.39	(0.09)***
ij level 2 belief strength											0.29	(0.08)***
ij level 3 belief strength											0.01	(0.10)
ji level 2 belief strength											−0.01	(0.09)
ji level 3 belief strength											−0.05	(0.10)
Constant	5.79	(0.33)***	6.43	(0.46)***	5.86	(0.38)***	6.89	(0.57)***	3.74	(0.81)***	3.01	(1.02)**
R^2	0.14		0.13		0.14		0.20		0.21		0.32	
Chi^2	18.91**		25.01**		19.07*		28.19**		57.64***		74.96***	
ji trust level												
ij same location	0.69	(0.21)***	0.66	(0.21)***	0.68	(0.21)***	0.62	(0.21)**	0.32	(0.21)+	0.24	(0.21)
ij same division	−0.11	(0.21)	−0.20	(0.21)	−0.09	(0.22)	−0.14	(0.22)	−0.09	(0.21)	−0.14	(0.20)
ij same gender	−0.34	(0.26)+	−0.51	(0.27)*	−0.35	(0.26)+	−0.58	(0.27)*	−0.32	(0.27)	−0.30	(0.26)
i male	−0.10	(0.25)	−0.16	(0.25)	−0.10	(0.25)	−0.15	(0.25)	−0.15	(0.24)	−0.04	(0.24)
i tenure	−0.02	(0.01)*	−0.02	(0.01)*	−0.03	(0.01)**	−0.02	(0.01)*	−0.03	(0.01)**	−0.03	(0.01)**
j male	0.15	(0.25)	0.11	(0.25)	0.15	(0.25)	0.10	(0.25)	0.14	(0.24)	−0.01	(0.23)
j tenure	0.01	(0.01)	0.02	(0.01)+	0.01	(0.01)	0.02	(0.01)+	0.03	(0.01)**	0.03	(0.01)**
i betweenness			−103.63	(64.94)+			−135.80	(78.71)*	−126.71	(76.36)*	−127.49	(76.35)*
j betweenness			−62.32	(62.57)			−82.07	(77.08)	−42.07	(76.82)	−18.20	(73.87)
i constraint					2.36	(1.87)	4.07	(2.24)*	3.91	(2.17)*	3.27	(2.06)+
j constraint					0.77	(1.89)	1.59	(2.29)	0.07	(2.25)	0.57	(2.16)
ij trust level									0.47	(0.09)***	0.42	(0.09)***
ij level 2 belief strength											−0.06	(0.09)
ij level 3 belief strength											0.00	(0.11)
ji level 2 belief strength											0.26	(0.09)**
ji level 3 belief strength											0.04	(0.11)
Constant	5.86	(0.33)***	6.46	(0.47)***	5.94	(0.38)***	6.90	(0.58)***	3.63	(0.84)**	2.75	(1.06)**
R^2	0.14		0.17		0.15		0.18		0.19		0.29	
Chi^2	18.61**		22.53**		20.84*		24.47**		53.68***		67.30***	

Note. $N = 113$.
$^+p < 0.10$; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

communication is truthful, but what is being communicated may not be "the whole truth." Trust neighborhoods can be used to represent referral cliques, often used to get the "inside story" on people and organizations that have a history of interactions within an industry. A referral clique is a clique of actors in which information flows are "trustworthy" in the technical senses of trust that we introduced earlier. Within a trust neighborhood (a "circle of trust"), sensitive information is likely to flow more reliably and accurately than outside of it. The epistemic approach we have taken to representing trust allows us to map the precise trust neighborhoods within a network, and therefore to make predictions about the reliable spread of accurate information within the broader network of contacts.

We define *security neighborhood* $N_S(G)$ as a fully linked subnetwork of G (clique) that shares trust in mutual competence and integrity (strong form). Security neighborhoods are high-trust cliques, and may be good representations for networks of field operatives in law enforcement scenarios, where authentication of communicated information is crucially important to the payoff to each actor, or conspiratorial networks, for example, a subgroup of a board of directors trying to oust the company's chief executive officer. The reason for calling such a network a security neighborhood is that it has (common) knowledge of itself *as* a

trust neighborhood, and thus possesses an important authentication mechanism for communication, which trust neighborhoods based on weak-form trust do not possess. If A and B belong to the same trust neighborhood, then C (a third party unknown to A but known to B) can interject herself in the communication between A and B and contribute information to A and B. If the fact that C is part of the trust neighborhood is not itself common knowledge, then the information that C contributes is not authenticated, in the sense that it is not possible for A to decide without consulting B whether or not to trust information coming from C. If, on the other hand, A and B are part of a security neighborhood, then C will immediately be recognized as an outsider. Common knowledge of clique membership is a key element in the process of authentication, which is itself important in subnetworks concerned with infiltration from the outside. A key property of a security neighborhood is thus that it necessarily has common knowledge of the fact that it is a security neighborhood, as well as of any fact that is relevant to it:

Proposition 1. *A security neighborhood is a common knowledge neighborhood.*

Proof. Consider a two-person clique (A, B), where ATB and BTA, both in the strong sense. For any P, it is the case that $Ak('P = \text{“}true\text{”} \leftrightarrow BaP)$ and $Bk(P = \text{“}true\text{”} \leftrightarrow AaP)$. For instance, let $P = \text{“}ATB.\text{”}$ Because P is true and $Ak('P = \text{“}true\text{”} \leftrightarrow BaP)$, $Ba(ATB)$ is true. Now, suppose P is true, but there is some level of almost common knowledge, n, at which it is not the case that $(AkBk)^nP$, i.e., $\sim (AkBk)^nP$. Then, at almost-common-knowledge level $n-1$, it cannot be the case that $(AkBk)^{n-1}P$ (to see this, let P^{n-1} represent the proposition “$(AkBk)^{n-1}P$,” which is true and together with ATB implies $(AkBk)^nP$). Therefore, if P is true, ATB, and BTA, then P is common knowledge.

Concatenating trustful/trusting relationships, we can define the most likely paths by which relevant and truthful information will flow in a social network as follows.

We define a *trust conduit* from actor A to actor J as a path A–B–$C \cdots J$, from A to J, passing through actors B–I such that A trusts B, B trusts C, C trusts $D, \ldots, I$ trusts J. Trust conduits can enable reliable knowledge flows in networks: information that comes from A will be trusted by B, information that comes from B will be trusted by C, and so forth, such that information flows credibly along a trust conduit. Trust conduits can represent knowledge pipes in organizations and markets, and thus to study the dynamics of the propagation of new information. They can also be used to affect a useful distinction between facts and rumors: facts are bits of information that have propagated along a trust conduit, whereas rumors are bits of information that have not. Because facts can be used to check rumors, trust conduits not only enable speedy propagation of useful relevant information, but also to provide checks and constraints on the propagation of rumors. Rumors should thus die out more rapidly in networks seeded with many trust conduits than those lacking them.

Because trust relations are not necessarily symmetric, the following distinction is useful: A *trust corridor* is a two-way trust conduit. A trust corridor is useful for representing reliable bidirectional knowledge flows within a network, thus increasing the degrees of freedom associated with any particular flow. If reliable knowledge can flow in both directions in a knowledge pipeline, rumor verification can proceed more efficiently, because any one of the actors along the path of the corridor can use both upstream and downstream actors for verification purposes. Trust corridors may be good representations for *expert networks* comprised of actors who can verify relevant rumors that come to the attention of any one of the actors in question. Trust corridors therefore can be seen to both accelerate the reliable transmission of facts and also to impede the promulgation of unverified or unverifiable rumors within a network. Figure 3 illustrates the trust neighborhood, conduits, and corridors graphically. Figure 4 graphs these epistemic regimes for the telecommunications firm we studied. If trust safeguards effective communication of reliable and accurate information, then one would expect that relevant bits of information will propagate relatively faster and more efficiently within trust neighborhoods and along trust and security conduits.

Special kinds of network relationships are required for the propagation of “sensitive” information—information that senders wish to be assured reaches

Figure 3 Trust Regimes

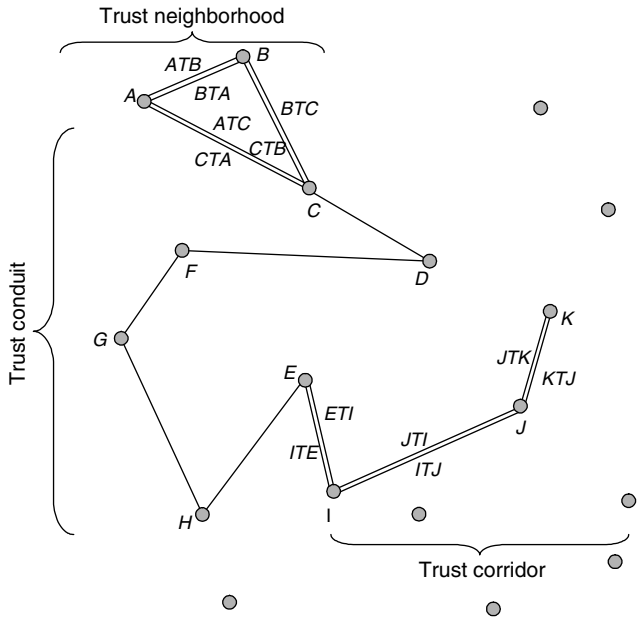
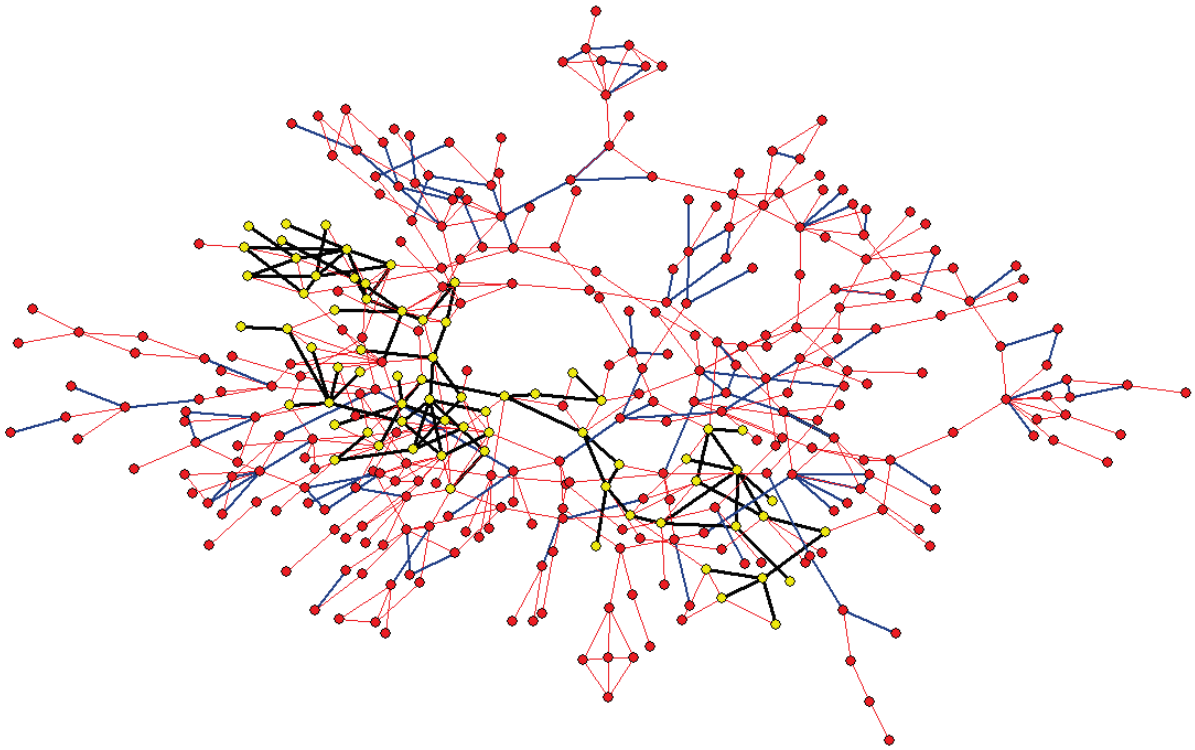


Figure 4 Trust Regimes Within a Large Telecommunications Company



Note. Yellow nodes and black arcs represent major trust neighborhoods, corridors, and conduits; red nodes and blue arcs represent secondary trust neighborhoods, corridors, and conduits.

all and only intended recipients. In such situations, actors' knowledge of the integrity of the trust conduit they access is required for the conduit to function as an information channel. The conditions for the existence of such conduits can be made precise by the use of the epistemic description language we developed above as follows.

A *security conduit* from actor A to actor J is a path A–B–$C\cdots J$, from A to J, passing through actors B–I such that A trusts B, B trusts C, C trusts $D,\ldots,I$ trusts J, and the conduit is common knowledge among $(A,\ldots,J)$. This feature renders them robust against infiltration, because each user of the communication channel embodied in the conduit can be authenticated by any of the actors comprising the conduit. Security conduits can be understood as authenticated knowledge pipes representative of operative actor networks (e.g., secret actor networks) in which rapid authentication of incoming messages can be performed quickly using actors' (common) knowledge of each others' membership in the conduit.

A *security corridor* is a two-way security conduit. It can be used to represent subnetworks in which (a) rumor verification can be performed more rapidly than in one-way knowledge-conducting structures, and (b) the probability of undesired "infiltration" is low because of the authentication mechanism proper to "strong-form" trust-based subnetworks. Security corridors may also be used to represent authenticated expert networks, which are characteristic, for instance, of robust social referral networks, which cannot easily be infiltrated, and which can be used for both fast rumor verification and the robust circulation of reliable knowledge.

Taken together, the epistemic building blocks we have introduced above allow researchers to study not only the impact of trust on coordination and mobilization, but also the impact of trust and trust-related structures (neighborhoods, conduits, corridors, security neighborhoods, corridors, conduits) on the propagation of information within a network. Seen through this lens, trust appears to provide a mechanism for the *informational superconductivity* of social networks and an explanation for the differential propagation of information within organizations and markets.

The epistemic approach to defining trust also allows us to probe into subtler aspects of the propagation of critical information in networks and, in particular, to make predictions about the relative conductance of a trust conduit (Nohria 2010) and the effects of perceptions about the relevance of a piece of information on the dynamics of that information.

Suppose Alina communicates "in confidence" a secret piece of information S to Belinda, whom she trusts. Belinda immediately turns around and lets Claire, whom Belinda trusts, in on S. Claire proceeds in a similar fashion, and within days S becomes shared knowledge within the organization (even though very few know that others know it). Has Belinda broken Alina's trust? If so, *where*? The epistemic definition of trust we introduced suggests a simple answer: Belinda broke Alina's trust in her when she did not tell Alina that Belinda told Claire what Alina had told Belinda in confidence (namely, S), because the proposition "I told Claire S" is relevant and true, and Belinda knows it to be relevant and true. However, if Belinda trusts Claire, then she can justify not telling Alina that now Claire knows S by reasoning that because Claire will keep the information to herself, S will not get any further than Claire within the organization, which is something that Alina would not consider relevant. Of course, if Claire reasons the same way about Belinda, S will freely propagate along, and *become* relevant as a result of the very way in which Belinda has defined relevance. The examples should highlight both the importance of a trust conduit in the propagation of "secrets" and the insidious effects of trust in promoting the very kinds of behavior that trust seems designed to counteract.

Discussion and Conclusion

We have argued for an epistemic and interactive definition of trust and for the importance of trust so defined to cooperative behavior in networks of human actors. Trust is a complex epistemic quantity that can be measured, and its impact on the propensity of human networks to comobilize and coordinate can be predicted. Various epistemic structures and information flow regimes that arise in organizational settings can be understood using the basic building blocks of trust in competence and trust in integrity. We showed how trust can be defined in terms of the epistemic states of networked agents, how that definition can be used to measure trust in a noncircular fashion, and how epistemically defined trust relationships can plausibly function as safeguards for coordination and information flow in networks. Because trust is defined precisely in terms of agent-level epistemic states, it is possible to both measure trust noncircularly using standard survey methods and also manipulate agent-level epistemic states and observe the effects of such manipulations on trust. Moreover, the proposed model also makes it possible to measure the effect of trust on the propagation of information, knowledge, and beliefs within a network. This latter direction also likely represents the most obvious of the methods and results of this paper, and would extend the contribution of this paper to an understanding of the mechanisms by which brokerage across structural holes bestows network structure advantage upon structural hole spanners.

Because trust is defined as a relationship between agent-level epistemic states, the epistemic description language we introduced enables researchers to study not only the effects of trust on higher-level epistemic states and on the flow of information in networks, but also the dynamics of *trust making and trust breaking* phenomena. The Marquis de La Rochefoucauld pointed out in his *Maximes* that it is impossible to distinguish between a person who behaves in a trustworthy fashion because that is the kind of person that he is and a person who behaves trustworthily for a (potentially very long) period of time to advantageously exploit the trust that he has slowly and meticulously built up (La Rochefoucauld 1665). Our analysis of trust allows for this kind of La Rochefoucauldian manipulation: If A trusts B, then A's trust of B creates opportunities for B to take advantage of A's trust, should she wish to do so. Indeed, in such a case, it is B's consistent and long-run refusal to take advantage of these opportunities in situations in which doing so has negligible costs that gives the trust relationship its unusual force. Of course, if a trusted agent (B) knows that the trusting agent (A) knows this, then the possibility that B will instrumentally manipulate A's trust will rear its head once again, which highlights the importance of the precise and accurate measurement of higher-level beliefs when interpreting behavior.

At the same time, trust is a delicate relationship because it only takes only one counterexample to the general-form statement "if P were true, then B would assert P or act as if P were true" to refute the proposition and therefore to undermine the trust that A has in B, whereas it takes an infinite and therefore infeasible number of confirmations to prove it. It may be (correctly) objected that this account of trust is too simple because it leaves out phenomena such as contrition and generosity, valuable under noisy conditions (Axelrod 1997). Considering the effects of epistemic structures in such cases is beyond the scope of this paper, but it is clear that higher-level epistemic states will be relevant to a comprehensive investigation of these phenomena. For instance, if A has reason to believe that B construes the interaction environment as noisy, then A can exploit the fact that B is likely to interpret A's trust-breaching actions as noise. If B does not know this, then he is more likely to forgive A than if B knows it, in which case B may react even more punitively than he would in a noise-free environment, to punish A's double breach of trust.

Such observations make it possible for us to explain the endogenous formation of structural holes in networks of actors (Burt 1992) insofar as we understand structural holes as "trust holes" (of integrity, competence, or both). Suppose A trusts in both the integrity and competence of B, and B trusts in both the integrity and competence of C. Then A trusts in the integrity and competence of C (in the strong sense), and the subnetwork A–B–C constitutes a security neighborhood. Suppose that, as per La Rochefoucauld, B has worked on obtaining A's trust over time with the intention of breaking it at an opportune moment, and that the opportune moment is the possibility of creating a trust hole between A and C that only B can bridge. The key to B's implementation of his intention is the realization that the knowledge operator ("knows p," or Kp) is dependent upon a domain of relevance of what is known to the aims of the knower(s). Thus, if C knows P ("cedar needles are green") but P is not relevant to A (an investment banker who has never seen a cedar tree), then the fact that C does not assert P cannot constitute a breach of A's trust of C; otherwise, the trust relationship condition would be too strong because it would require the parties to constantly assert propositions that are true but irrelevant. Relevance is thus a condition that can be made as precise as the rest of the analysis so far: P is relevant to C, for instance, iff it is the case that C's knowledge of P will cause C to change her planned course of action in a particular case.

Armed with this analysis, B, our trust breaker, can now exploit small and potential transient asymmetries of information or insight to break off A's trust in C and therefore create the A–C trust hole that only he, B, can bridge. In particular, to create the hole, B will set up a situation in which (a) C knows P but believes that P is not relevant to A, (b) P is relevant to A, (c) A knows C knows P, and (d) C does not assert P, which will be deemed by A to be in contradiction of the proposition "if P were true, A would assert P," which underlies her trust in C. Using this blueprint for the undermining of trust, B can look for situations that allow him to exploit asymmetries of understanding between A and C, wherein A and C attach different degrees of relevance to the same proposition, P, which both, nonetheless, consider to be true. Upon finding such a proposition, B must then ascertain that A knows that C knows P and rely on the mismatch in ascribed degrees of relevance to cause the breach of trust.

For instance, telecommunications industries (such as WiFi, based on the IEEE 802.11(a) standard, and WiMax, based on the IEEE 802.16(d–e) standards) rely on large, established chip makers like Intel that create system developer ecosystems around them that embed their chips into products. A system manufacturer with private information about a chip's limitations can cultivate Intel's trust by informing it of the limitations of the chip and of new market opportunities. And it can undermine Intel's trust in its other partners (such software developers that target their applications to a chip's capabilities) by communicating information that it privately knows (from other suppliers) to be relevant to Intel but also knows the partners will not themselves communicate in a timely fashion back to Intel because, for example, of their less intense interaction with the chip company. Thus, interaction intensity can be used to seed breaches of trust based on asymmetries of understanding.

We have, in summary, provided a description language for describing trust in social networks that is versatile enough for modeling both the role of trust in the dynamics of information and the role of information sharing and hoarding strategies in the formation and dissolution of trust.

Acknowledgments

The authors gratefully acknowledge the support of the Marcel Desautels Centre for Integrative Thinking at the Rotman School of Management for past and ongoing support of this work. They thank Tim Rowley for his help in securing the empirical data herein, and Diederik Van Liere and Laurina Zhang for their help coding them. They also thank Michel Anteby, Julie Battilana, Roy Chua, Jack Gabarro, Ranjay Gulati, Rakesh Khurana, Paul Lawrence, Nitin Nohria, Jeffrey Polzer, Lakshmi Ramarajan, Michael Tushman, and other participants in a Harvard Business School Organizational Behavior Area seminar on the topic of this paper for their insight and feedback. They are grateful to Jesper Sørensen, Wallace Hopp, Ron Burt, and two anonymous *Management Science* reviewers for their criticisms and helpful suggestions. All errors remain the authors' responsibility.

References

Arrow, K. J. 1974. *The Limits of Organization*. Norton, New York.

Aumann, R. J. 2001. Game theory. *New Palgrave Dictionary of Economics*. Blackwell, New York.

Aumann, R. J., A. Brandenburger. 1995. Epistemic conditions for nash equilibrium. *Econometrica* **63**(5) 1161–1180.

Axelrod, R. 1997. *The Complexity of Cooperation: Actor-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, NJ.

Ayres, I., B. Nalebuff. 1997. Common knowledge as a barrier to negotiation. *UCLA Law Rev.* **44** 1631–1659.

Barber, B. 1983. *The Logic and Limits of Trust*. Rutgers University Press, New Brunswick, NJ.

Barney, J., M. Hansen. 1994. Trustworthiness as a source of competitive advantage. *Strategic Management J.* **15**(S1) 175–190.

Belsley, E. K., R. E. Welch. 1980. *Regression Diagnostics*. Wiley, New York.

Blau, P. 1964. *Exchange and Power in Social Life*. Wiley, New York.

Bourdieu, P., L. Wacquant. 1992. *An Invitation to Reflexive Sociology*. University of Chicago Press, Chicago.

Brandenburger, A. 1992. Knowledge and equilibrium in games. *J. Econom. Perspectives* **6**(4) 83–101.

Burt, R. S. 1992. *Structural Holes*. Harvard University Press, Cambridge, MA.

Burt, R. S., M. Knez. 1995. Kinds of third-party effects on trust. *Rationality Soc.* **7**(3) 255–292.

Butler, J. K. 1991. Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *J. Management* **17**(3) 643–663.

Chwe, M. S.-Y. 1999. Structure and strategy in collective action. *Amer. J. Sociol.* **105**(1) 128–156.

Chwe, M. S.-Y. 2000. Communication and coordination in social networks. *Rev. Econom. Stud.* **67**(1) 1–16.

Coleman, J. S. 1988. Social capital in the creation of human capital. *Amer. J. Sociol.* **94** S95–S120.

Coleman, J. S. 1990. *Foundations of Social Theory*. Harvard University Press, Cambridge, MA.

de Tocqueville, A. 1864. *Democracy in America*. Sever and Francis, Cambridge, UK.

Durkheim, E. 1933. *The Division of Labor in Society*. MacMillan, New York.

Freeman, L. C. 1977. A set of measures of centrality based on betweenness. *Sociometry* **40**(1) 35–41.

Friedland, N. 1990. Attribution of control as a determinant of cooperation in exchange interactions. *J. Appl. Soc. Psych.* **20**(4) 303–320.

Fukuyama, F. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, New York.

Gambetta, D. 1988. *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell, New York.

Gettier, E. L. 1963. Is justified true belief knowledge? *Analysis* **23**(6) 121–123.

Golembiewski, R. T., M. McConkie. 1975. The centrality of interpersonal trust in group processes. C. L. Cooper, ed. *Theories of Group Processes*. John Wiley & Sons, New York, 131–185.

Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *Amer. J. Sociol.* **91**(3) 481–510.

Greene, W. 2000. *Econometric Analysis*. Prentice Hall, New York.

Hill, C. W. L. 1990. Cooperation, opportunism, and the invisible hand: Implications for transaction cost theory. *Acad. Management Rev.* **15**(3) 500–513.

Hosmer, L. T. 1995. Trust: The connecting link between organizational theory and philosophical ethics. *Acad. Management Rev.* **20**(2) 379–403.

Kramer, R. 1999. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Rev. Psych.* **50** 569–598.

Kreps, D. M. 1990. Corporate culture and economic theory. J. E. Alt, K. A. Shepsle, eds. *Perspectives on Positive Political Economy*. Cambridge University Press, New York, 90–143.

La Rochefoucauld, F. 1665. *Reflexions ou Sentences et Maximes Morales*. Penguin, New York.

Lewis, J. D., A. Weigert. 1985. Trust as social reality. *Soc. Forces* **63** 967–985.

Mayer, R. C., J. H. Davis, F. D. Schoorman. 1995. An integrative model of organizational trust. *Acad. Management Rev.* **20**(3) 709–734.

Moldoveanu, M. C., H. Stevenson. 1998. Ethical universals in practice: An analysis of five principles. *J. Socio-Econom.* **27**(6) 721–752.

Moldoveanu, M. C., J. A. C. Baum, T. J. Rowley. 2003. Information regimes, information strategies and the evolution of interfirm network topologies. F. J. Yammarino, F. Dansereau, eds. *Research in Multi-Level Issues*, Vol. 2. Elsevier, Oxford, UK, 221–264.

Nohria, N. 2010. Personal communication.

Nov, O., S. Rafaeli. 2009. Measuring the premium on common knowledge in computer-mediated coordination problems. *Comput. Human Behav.* **25**(1) 171–174.

Rubinstein, A. 1986. Finite automata play the repeated prisoner's dilemma. *J. Econom. Theory* **39**(1) 83–96.

Schelling, T. C. 1978. *Micromotives and Macrobehavior*. Norton, New York.

Weber, M. 1992. *The Protestant Ethic and the Spirit of Capitalism*. trans. T. Parsons, London, Routledge.

Williamson, O. E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press, New York.

Williamson, O. E. 1985. *The Economic Institutions of Capitalism*. Free Press, New York.