Analyse & Kritik 28/2004 (© Lucius & Lucius, Stuttgart) p. xxx–xxx

Coye Cheshire/Karen S. Cook

The Emergence of Trust Networks under Uncertainty – Implications for Internet Interactions

Abstract: Computer-mediated interaction on the Internet provides new opportunities to examine the links between reputation, risk, and the development of trust between individuals who engage in various types of exchange. In this article, we comment on the application of experimental sociological research to different types of computer-mediated social interactions, with particular attention to the emergence of what we call '*trust networks*' (networks of those one views as trustworthy). Drawing on the existing categorization systems that have been used in experimental social psychology, we relate the various forms of computer-mediated exchange to selected findings from experimental research. We develop a simple typology based on the intersection of random versus fixed-partner social dilemma games, and repeated versus one-shot interaction situations. By crossing these two types of social dilemma games and two types of interaction situations, we show that many forms of Internet exchange can be categorized effectively into four mutually exclusive categories. The resulting classification system helps to integrate the existing research on trust in experimental social psychology with the emerging field of computer-mediated exchange.

1. Introduction

We investigate recent experimental work on social exchange networks, uncertainty, commitment and trust to analyze emerging forms of social interaction that are mediated by computers including systems like the Internet that link people across the globe. In particular we examine different types of exchange networks and the role of reputations and trust in these networks, commenting on the application of this research to various types of computer-mediated social interaction. We focus attention on the emergence of what we call 'trust networks' (Cook/Rice/Gerbasi forthcoming) to resolve some of the difficulties involved in anonymous exchanges of different types of goods and services. Trust networks are the egocentric networks of those one views as trustworthy.

Computer-mediated exchange allows individuals to exchange with various degrees of anonymity through electronic networks. Since anonymity creates uncertainty and risk, resolving this uncertainty and developing ways to avoid risk are central to creating ongoing exchange environments. Although reputation systems are shown to be a well-supported solution to the inherent uncertainty and risk in this type of exchange situation, we conclude that reputation systems are only one mechanism for attempting to secure trustworthiness and reliability.

We comment on other mechanisms that have been developed in efforts to ensure reliability, particularly when reputation information is not available.

To understand the emergence of trust networks in computer-mediated communication settings such as those on the Internet, we develop a typology based on the intersection of random versus fixed-partner social dilemma games, and repeated versus one-shot interaction situations. By crossing these two types of social dilemma games and two types of interaction situations, we show that many forms of Internet exchange can be categorized effectively into four mutually exclusive categories. Since similar categorization systems have been used in experimental social psychology, we relate the various forms of computer-mediated exchange to the selected findings from experimental research. We begin with a review of the pertinent findings from experimental social psychology on uncertainty and commitment in exchange relations.

2. Uncertainty and Commitment in Exchange Networks

Experimental research on exchange relations in social networks indicates that under varying degrees of uncertainty actors begin to form commitments to their exchange partners when they believe there is risk in the environment, typically the risk of exploitation (i.e., Cook/Emerson 1978; Kollock 1999a; Yamagishi/Matsuda 2003; and Rice unpublished). This is a form of interpersonal commitment even though the actors may know very little about each other except their own history of exchange revealed over time through their interactions. This behavioral history and its veracity may be the most important source of reputation in networks of mediated interaction.

Uncertainty concerning the quality of the goods at stake is one reason for commitment formation. When a buyer finds a supplier who is reliably trustworthy in terms of providing goods of high quality the buyer may try to enter a committed relationship with that particular seller. As Kollock's (1994) experimental research suggests those who locate partners who are trustworthy tend to form committed relations. In markets where quality can be determined more easily by the buyer (as is the case with the rice market, unlike the rubber market) actors are less likely to form commitments and thus they remain open to new exchange partners. Cook/Rice/Gerbasi (forthcoming) explore the consequences of commitment formation and the emergence of trust networks for the transition to open market economies.

In Kollock's (1994) study of opportunistic uncertainty and commitment, actors exchanged goods in two different environments. In one environment (low uncertainty) the true value of the goods being exchanged was known, while in the other (high uncertainty) environment the true value of goods was withheld until the end of the negotiations. Actors had a greater tendency to form commitments in the higher uncertainty environment. They chose to continue to transact with the partners who had established their trustworthiness by not misrepresenting the value of their goods rather than to engage in potentially more profitable exchanges with new, untested partners.

Yamagishi/Cook/Watabe (1998) further explored the connection between uncertainty and commitment. In their experiment, actors were faced with the decision of remaining with a given partner or entering a pool of unknown potential partners. They employed several modifications of this basic experimental design, but in each instance the expected value of exchange outside the existing relation was higher than the returns from the current relation. They found that actors were willing to incur sizeable opportunity costs to reduce the risks associated with opportunism. Moreover, they found that uncertainty in either the form of an uncertain probability of loss or an unknown size of loss promoted commitment between exchange partners.

Given high levels of uncertainty in most systems of Internet trade, commitments are likely to form in markets for certain types of goods and services. The most problematic trades for buyers are the one-shot interactions in which either there is no need for a repeat trade in the near future or it is unlikely that the actors involved would ever engage in repeat trade for a number of reasons (unrelated to the type of good or service involved).

3. The Emergence of Trust Networks under Uncertainty

Our research indicates that commitments tend to form under uncertainty and that trust is higher among those who develop such behavioral commitments. However, it is also clear in our research that when trust is measured as an independent variable (as a predisposing factor) high trusters and low trusters tend to behave differently in the same environments. Under high uncertainty it is the low trusters who are the quickest to form interpersonal commitments. High trusters are more likely to be risk takers and to remain in the "market" for new partners longer.

In our most recent cross-cultural study of trust formation in the U.S. and Japan, we find that cooperation among individuals is significantly higher when individuals are given the opportunity to have repeat interactions with the same partners instead of being randomly paired with partners (Cook et al. forthcoming). In these experiments, individuals are given the ability to 1) control how much to entrust to a partner and, 2) decide whether to cooperate with the partner by returning (or not returning) the amount that was entrusted to them. When individuals only make a decision about whether to return coins or not (and the amount is decided by the computer), we consider this to be a basic prisoner's dilemma. When individuals have *both* of these choices, however, we call this a PD/R game, or, a *prisoner's dilemma with risk*. We found that partoo95TD tTf125.72ooo95T52l(Tf17TDucnoall)-38353lenindivtmage 4o950-110-1283(is)-383(di950-11.95

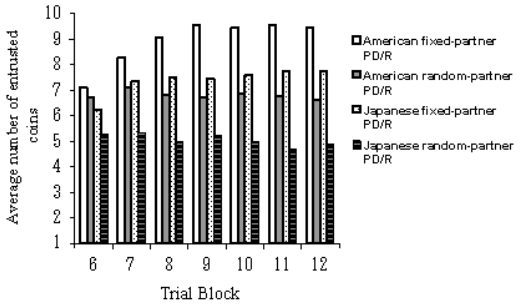

Figure 1: Average number of entrusted coins across trial blocks in fixed versus random partner exchange using the prisoner's dilemma with risk (PD/R) in the U.S. and Japan (Cook et al. forthcoming).

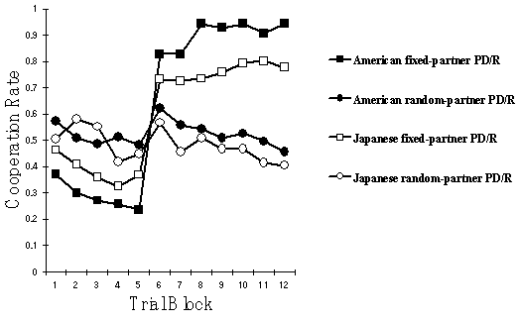

Figure 2: Average cooperation rates (proportion of coins returned) across trial blocks for both American and Japanese participants using the prisoner's dilemma with risk (PD/R) with fixed and random partners (Cook et al. forthcoming).

control this amount. We found that allowing individuals to signal their trustworthiness seems to have a positive effect on cooperation, but this positive effect is minor and short-lived. In fact, the effect is not significantly stronger than in interactions between random partners who cannot signal each other (Cook et al. forthcoming). Furthermore, we found that there is no cross-cultural difference in levels of cooperation between American and the Japanese participants when they participate in random-partner exchange.

In real-world situations in which trade partners are often initially unknown, having some system for establishing their trustworthiness is a primary concern, especially in settings in which it would be easy to cheat. Institutional backing, law and other devices for managing distrust or the breach of trust are crucial in the development of systems that manage trades, auctions and other forms of interaction involving the transfer of goods, services and money reliably.

Computer-mediated exchange on the Internet, for example, is an environment in which trading and auctions have flourished—despite concerns about anonymity and the selection of trade partners at random initially. One reason for the success of many Internet exchange systems is that these systems have either adopted reputation systems, or such systems have evolved out of necessity. In the following sections, we discuss the emergence of reputation systems and the conditions under which different types of reputation systems are likely to work for mediated interactions (i.e. Internet trades). We also discuss alternatives to reputation systems, especially in exchange situations in which it is impractical or impossible to obtain reputation information.

The Role of Reputation Systems in Exchanges

Yamagishi/Matsuda (2003) demonstrate that the role of reputations varies depending upon whether the social system in which the actors are embedded is closed or open. In closed communities reputations can be effective because negative reputations can lead to exclusion (Cook/Hardin 2001). In more open societies negative reputations are less effective because they are limited in the extent to which they are transferred to all those in the network (or social system) since information flows only across those actors that are linked directly or indirectly. More importantly, actors can alter their identities in ways that make it easier to reenter trade networks without being recognized as having had a negative reputation. The focus of Yamagishi/Matsuda's (2003) experimental study is an Internet trading network in which both the level of honesty in the trades and the price can be tracked. Positive - reputation mechanisms are more effective than negative - reputation mechanisms in open networks. In both cases (positive and negative reputation conditions) actors are rated on their honesty and thus accumulate reputation points that are published in the network after each transaction.

In this research reputation mechanisms are investigated by Yamagishi and Matsuda (2003) as solutions to the "lemons" problem in Internet markets, that is, the tendency for the goods on the market to be low quality especially when there is asymmetric information between buyers and sellers. Typically, the lemons market (Akerlof 1970) emerges when the buyers have much less valid or reliable information about the quality of the goods on the market than do the sellers (as is frequently the case in the used car market). Buyers must rely on the reputation of the sellers to determine how much confidence they have in the seller (i.e. how trustworthy the seller is). In this world of trade Yamagishi and Matsuda (2003) demonstrate that the lemons problem is exacerbated when actors can change their identities and reenter the market with a new identity (thus canceling their negative reputation). They also demonstrate that positive reputations affect behavior differently since actors develop an investment in their reputations and want to protect them. Traders lose their positive reputation and their investment in it if they alter their identities. In fact, the negative reputation mechanism designed to reveal dishonest traders is particularly vulnerable to identity changes, whereas the positive reputation mechanism designed to identify honest traders is not so vulnerable to identity changes. The results

concerning the quality of goods produced in settings using these two different reputation mechanisms (with and without the possibility of changing identities in the exchange network) are presented in Figure 3 (see Yamagishi forthcoming for details). The findings clearly demonstrate the superiority of positive reputation mechanisms especially when identities can be easily changed as is typically the case in Internet trading.

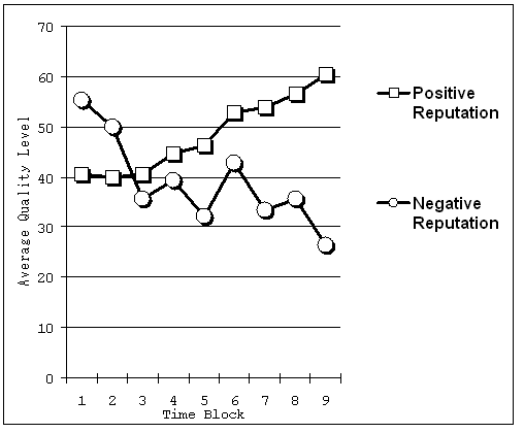

Figure 3: Average quality level of goods exchanged in positive versus negative reputation systems (Yamagishi/ Matsuda 2003)

To test the prediction that positive reputation devices are more effective than negative reputation mechanisms, Yamagishi/Matsuda (2003) conducted one experiment in which the quality level of the goods could be compared between exchange systems involving only positive or only negative reputations. The findings show that there is a significantly higher quality level, at least initially, when the negative reputation mechanism is used. Yet, as negative reputations grow, so do the tendencies to create new identities to shed the old negative reputation and begin anew. The positive reputation device proved to be better for producing higher quality goods in the long run when compared to a purely negative reputation system. Individuals became invested in their positive reputations.

The negative reputation device may work best under the condition that someone who develops a negative reputation can be effectively excluded from trade by the group acting as a whole. This requires a closed network with information sharing[1]. Norms of exclusion work only in groups and are not effective in open networks. They function to eliminate those who have "cheated" from the system of trade (Hardin 1995; Cook/Hardin 2001).

Kollock (1999a) views the auction houses that have been created for trade on the Internet as a laboratory for the study of the management of risk when there is little or no access to external enforcement mechanisms (the typical way we

[1] See also Grief (1993) for a discussion of the difference between closed associations and open trade networks among the Maghribi.

manage risk). He studies the emergence of endogenous solutions to the problems of risky trade in cases in which no guarantees, warranties, or other third-party enforcement mechanisms exist. The reputation systems that emerge to manage this risk are the primary focus of his research.

Yamagishi/Matsuda (2003) explore the differences documented by Kollock between negative and positive reputation systems. Kollock (1999a, 111) notes that "a particularly disturbing strategy for fraud...in a number of online markets is the person who works hard at establishing a trustworthy reputation, and then sets up a whole series of significant trades and defaults on all of them, disappearing into a new identity after the fact." In a quantitative study of auctions on eBay, Kollock (1999a) provides evidence that at least for some high-value goods, the price buyers paid was positively and significantly affected by the seller's reputation.

In a simulation study of the effects of positive reputation systems that are similar in character to the positive reputations studied experimentally by Yamagishi/Matsuda (2003), Whitmeyer (2000,196) finds that the effects of different types of positive reputation systems depend to a large extent on the proportion of cooperators (as opposed to non-cooperators) in the population. He examines the effects of reputation systems on general confidence gains in the society. If the proportion of non-cooperators is low, any type of positive reputation system will work because the non-cooperators will be more easily detected, especially if it is hard to obtain a positive reputation.[2]. If the proportion of non-cooperators is high, then a tough reputation system will mean lost opportunities for cooperation because some potential cooperators will not be detected.

Alternatives to Reputation Systems for Securing Trustworthiness

In some exchange systems, it is not possible to establish reputation information, or it would at least not be feasible to do so. If an exchange system involves many one-shot interactions between random partners, it may not be possible to build a reputation (or the obstacles associated with regulating such a system may be too great). In such cases, there may be alternative types of mechanisms that emerge to provide *security*.

The extent to which there is institutional backing for failed trust is what matters most in these environments independent of the quality or veracity of the reputation information involved in the decision of whom to trade with (or buy from), especially when such information is unreliable or impossible to obtain. For example, the Maghribi traders of the 11^{th} century conducted trades across the Mediterranean despite the high level of uncertainty about their trading partners and, ultimately, the quality of the goods being traded (Grief 1989; 1993). Despite the incentives to cheat each other in these trades, the Maghribi were able to sustain successful trades. Grief (1989) explains this result by emphasizing the fact that the Maghribi traders were able to use closed networks that eliminated traders from their coalitions if they tried to cheat other traders. Community

[2] Whitmeyer varied the degree to which it was easy or hard to get a positive reputation in his simulation.

responsibility norms, as a kind of monitoring and policing device, developed as a way to secure trustworthiness within a circumscribed set of actors.

The social contexts in which exchanges are embedded are thus crucial for understanding how trust and cooperation can develop, especially in the absence of reputation information. Social context refers to the nature of the situations in which social interactions take place (Cook/Cooper 2003). Factors such as group size and network density, for example, can have an independent effect on the development of trust. Systems that are too large can reduce the effect that trust has on cooperative behavior (Sato 1988). Furthermore, the very emergence of trust and cooperation may depend on locally embedded "neighborhoods" that act as the source of cooperative and trusting behavior (Macy/Skvoretz 1998).

Another important factor that affects cooperation and the emergence of trust is the ability to communicate with partners. The ability to communicate with one's partner before playing an experimental social dilemma game has a strong effect on specific trust of one's partner (Mori 1996). Other social context variables have been shown to affect levels of trust and cooperation, including time pressure, the presence of third parties, and culture[3].

4. Implications for the Internet and other Forms of Computer Mediated Interaction

Trust networks can take several different forms in computer-mediated interaction and on the Internet. Although the Internet is certainly the largest computer-mediated network, we argue that the processes by which trustworthiness is established in such networks can be informed by the experimental work on trust (e.g., Dasgupta 1988; Kreps 1990; Snijders 1996; Berg/Dickhaut/McCabe 1995; Yamagishi/Kakiuchi 2000). The experimental work in this tradition uses social dilemma games (such as the classic prisoner's dilemma) as tools for understanding the processes by which trust and trustworthiness are established under conditions of risk and uncertainty.

A simple social dilemma game such as the prisoner's dilemma can be carried out as a one-shot game (in which the participants interact only one time), or as a repeated/iterated game (in which the participants interact several times). In addition, an iterated social dilemma game may have fixed partners (i.e, the same partner on every turn), or random partners (a new, random partner on every turn). Since one-shot games only involve one interaction, the fixed/random partner distinction does not apply. Thus, the intersection of these two dimensions produces a 4-cell matrix (with one empty cell) which is useful for categorizing the various types of interaction situations found in the Internet environment. In the following sections, we describe these different types of interaction situations and some of the issues related to establishing trustworthiness under varying con-

[3] For a more complete review of experimental studies of social context and trust, as well as a review of other experimental research on trust and cooperation, see Cook/Cooper (2003).

ditions. Table 1 includes examples of Internet interactions in each of the three main cells of our typology[4].

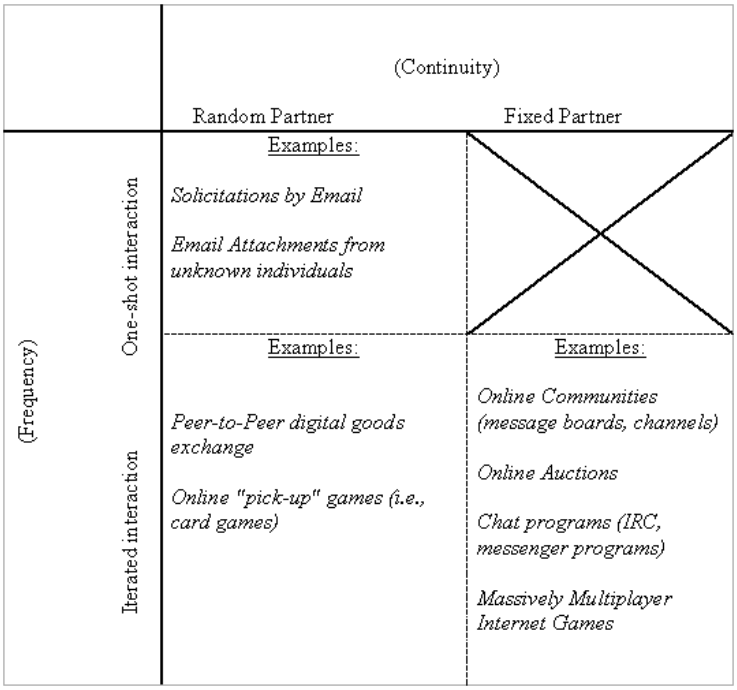

Table 1: Examples of types of Internet situations with random and fixed partners

One-Shot Interactions with Random Partners

In an experiment involving a one-shot situation with random partners, individuals must simultaneously make decisions whether to cooperate. Because the interaction only occurs once (or there is only one trial), both parties can either cooperate or not, but trust would rarely, if ever, emerge since there would be no repeat interactions. An act of cooperation in a one-shot situation can be viewed as involving taking a risk, rather than as an act of "trust" (see Hardin 2003). In a one-shot game there is no ability to send a signal to the partner by cooperating initially on several trials to signal trustworthiness. In computer-mediated communication, some types of Email interactions take this one-shot, random partner form.

[4] We view this typology as a useful tool for discussing the many types of interactions that take place on the Internet, yet we acknowledge that many of the examples in our typology can, and often do, take on different forms of interaction depending on the situation. As a result, the examples are not necessarily *always* one kind of interaction situation, nor do they *always* have fixed or random partners. It is precisely because of the malleability of interaction situations on the Internet that we find this kind of typology useful for describing any given situation.

For example, suppose a professor receives an email from an unknown person requesting an unpublished draft of a current research paper. The professor does not know the person, so how does she know if the person is not going to plagiarize the paper, or quote it against the wishes of the author? A different example of this sort of uncertainty exists when an individual receives an email from an unknown person with an attached file. How does the recipient know that the attachment is not a computer virus? In both situations, social context cues (Sproull/Kiesler 1986) such as the email address of the sender can become essential to establishing trustworthiness. If we assume that the recipient does not know the sender, then seemingly trivial things like the domain of the sender (e.g., .edu, .com), the lack or presence of a personalized address in the email or the tone of the message can have a profound influence on the response of the recipient. In the absence *of personal reputation information*, such cues act as a proxy for reputation information when attempting to establish the trustworthiness of one's partner (see also Gambetta/Hamill forthcoming).

Iterated Interactions with Random Partners

Computer-mediated interactions on the Internet that involve iterated interactions give participants multiple chances to cooperate with their partners. However, if the interactions are always with new partners sometimes selected at random, then it is not possible to establish personal reputation information because the individuals do not have previous knowledge of the level of the other's willingness to cooperate. In other words, in such situations individuals do not have a "shadow of the future" (Axelrod 1984) or any past that helps them determine what to expect from future interactions.

Kollock (1999b) argues that reciprocity in online communication is encouraged when four conditions are met: ongoing interaction, identity persistence, knowledge of previous interactions, and strong group boundaries. While individuals may occasionally be purely altruistic, Kollock (1999b) argues that this is not a necessary assumption. In particular, some individuals may just want to cooperate out of a sense of self-efficacy, an expectation of future rewards, or as a way to improve their own reputation. However, when interactions are mostly random (i.e., identity is not persistent), then the last two explanatory factors (i.e., to obtain future rewards and to attain or produce a reputation) do not apply because there is no way to identify oneself to receive rewards from others and to build a positive reputation.

On the Internet, one type of iterated interaction with random partners occurs in peer-to-peer exchange systems. In many popular peer-to-peer systems, individuals decide whether to share *digital information goods*, or just take such goods from others without sharing in return (Kollock 1999b). These *digital information goods* might take the form of digital media such as music or movies, or digital copies of written works such as electronic books or journal articles. Digital goods challenge many for-granted assumptions about the exchange of goods because digital goods can be copied from one person to another (i.e., the original owner does not lose value from the good during the exchange) and many people can benefit from the digital good without decreasing its value. Kollock

(1999b) argues that digital goods produced by a single person on the Internet can become instant public goods because so many can benefit from them. These goods are easy to provide partially because they do not require coordination costs that usually prevent individuals from producing a public good on their own.

The exchange of digital goods in peer-to-peer systems can be represented as a simple social dilemma in which an act of sharing digital goods is defined as "cooperation" and taking digital goods without sharing is defined as "defection". Individuals can be thought of as *random* partners because a given provider is determined by what a user is looking for in the peer-to-peer network. For example, if an individual wants to find a specific song title, when she searches for this title she is given a list of individuals who have that item and are currently sharing it. Because there are often hundreds of thousands of users, and there is no specific reputation information available, each contributor can be thought of as a random partner.

When an individual decides to retrieve a digital good from someone else, how does she know that the item will be what the provider says it will be? Furthermore, how can the person retrieving the digital good know that the item is not a computer virus or just bad data? In other words, just because an individual shares these goods, does not mean that she is trustworthy by default. There is some risk involved in downloading a digital good, such as a popular song, but the risk may be very small (i.e., involving only the time associated with downloading the file in question).

As with many types of real-world exchange situations, a clear way to establish trustworthiness in an iterated set of online interactions is through reputation information. As Xiong and Liu (2003) argue, trustworthiness in peer-to-peer systems can be determined if a transaction-based feedback system is used. In their case, a 5-point trust model is used to create reputation scores that help establish trustworthiness. While such reputation systems are desirable, *they are also not always practical and cannot be used in systems in which users are anonymous, or can easily change their identities* (as in many existing peer-to-peer systems). Thus, assuming that reputation systems cannot be used in iterated interactions among random partners; individuals must look for other devices for ensuring trustworthiness or reliability.

One way individuals infer the trustworthiness of providers without reputational information is by observing how many *other* people are also retrieving the digital good from that same provider. Since many peer-to-peer systems (i.e., Kazaa, Morpheus, and the original Napster) allow users to see how many other anonymous individuals are retrieving digital goods from any given provider, each individual can potentially infer the trustworthiness of the provider from the actions of others. Such information provides a kind of "virtual reputation" even if it is not specific enough. Of course, this may essentially be a conformity issue, because individuals may assume that if everyone else "trusts" the contributor or finds her reliable, the good is probably legitimate. In the absence of personal reputation information, social conformity can end up providing a proxy for true reputation information. This type of conformity may represent a kind of *herd*

behavior (sometimes seen in investment decisions), when a potentially infinite set of individuals make a similar decision (Smith/Sørensen 2000). Individuals who act in this manner do not necessarily ignore personal information. Instead, the behavior of other individuals *may be the only information that is available to the individual.* If an individual assumes that others may have more information than she does, then such "herd" behavior is rational.

Another way that individuals can infer the trustworthiness of a potential partner is through a third party that has already become established as trustworthy. A contemporary example would be when well-known and established companies set up online stores for the buying and selling of used items. Amazon.com, for example, allows individuals to buy and sell used goods directly from their website—yet Amazon.com does not handle the inventory or the actual exchange of used goods. Instead, Amazon.com acts as a reliable or trusted third party intermediary that markets items for sale, and handles the collection of payments.

Stefik (1999) argues that 'trusted systems' might be a solution to the problem of establishing confidence in digital goods and media. In this case, a trusted system might work like a bonded carrier, insuring the good and thereby reducing the risk that individuals face when exchanging digital goods (Lessig 2000). The key issue is that instead of the reputation of the individuals who are involved in a given exchange, it is *the reputation of the third party* that establishes trustworthiness or reliability. Thus, the risk associated with trusting other individuals is significantly reduced if some third party can either vouch for an individual, or possibly even underwrite or guarantee the transaction. Credit card markets work in this way in the consumer arena.

Iterated Interactions with Fixed (or Consistent) Partners

Iterated interactions among fixed (or consistent) partners on the Internet make up much of the existing research on Internet interactions and trust. Online auctions (Kollock 1999a; Yamagishi/Matsuda 2003) and virtual communities (Rheingold 1993; Curtis 1991; Abdul-Rahman/Hailes 2000) are characterized by repeated interactions with fixed partners. In such cases, fixed partners have stable identities that allow individuals to have repeated interactions that can be compared to previous interactions. As a result, having reputation information is a clear advantage for individuals who want to determine the trustworthiness of partners in such systems (Kollock 1999a; Abdul-Rahman/Hailes 2000; Yamagishi/Matsuda 2003).

An example of evolving virtual communities with iterated interactions among fixed (or consistent) partners is extremely large multiplayer games on the Internet. These games involve tens of thousands of players in a persistent online world. Players fight, trade, talk, and share experiences in these virtual online worlds. Raph Koster, a game developer for several of the world's largest multiplayer games argues that game developers should, "make sure that players have a reasonable expectation of future interaction. This means persistence of identity and limited mobility" (Koster 2003). The rationale is that players *need* reputation systems so that they know who they can and cannot trust.

Because computer-mediated exchanges take place without the benefit of face-to-face communication, it becomes especially difficult for real people to know how how much to trust others through digital avatars (i.e., an icon or representation of a real person in a shared virtual world). Citing Yamagishi/Matsuda's (2003) research on positive and negative reputation systems, Koster (2003) argues that these massively multiplayer games would be better served with positive reputation systems because, unlike negative reputation systems, they promote the persistence of identities.

Another example of Internet exchange with partners who are fixed (or who are at least *assumed* to be fixed), are Usenet newsgroups. The Usenet describes a system of organized newsgroups, or bulletin boards, that are available to almost anyone with Internet access. The Usenet does not provide assurance of reliability or authenticity—anyone with an account can post to a newsgroup under any name she chooses. The stability of one's identity is often a key part of ongoing interactions in these newsgroups. It is argued that stable identities establish a sense of community (Beniger 1987), group affiliation and support (Sproull/Kiesler 1991; Wellman/Gulia 1999), as well as the basic motivation to actively participate in ongoing conversations (Donath 1999).

The unregulated nature of Usenet newsgroups, however, also opens the possibility of identity switching and deception. As Donath (1999) describes in detail, a Usenet letter posted to any given newsgroup is filled with opportunities for deception and identity fabrication. A reader must take in several clues, including the email address of the sender, the content of the message, the language and tone of the message, and the signature of the sender. Is a given message *really* from a doctor if a message about medical advice is signed "Dr. Smith"? Is it more or less credible if the supposed doctor's email address is from an official sounding domain? Unfortunately, there is little way to be sure that any message is reliable or legitimate. For example, the account name in the header or letterhead can be fake, identity claims can be completely false, social cues can be deliberately misleading or faked (see Donath 1999, 44; Gambetta/Hamill forthcoming).

Although the kinds of Internet interaction situations described above can be viewed as repeated interactions with fixed partners, it is important to note that there are certainly exceptions to all of these examples. In fact, there is at least some research that shows that many types of Internet interaction situations are not always as 'iterated' as we might initially think. In a study on the quality of online social relationships, Cummings, Butler and Kraut (2002) show that some kinds of virtual communities such as newsgroups often have very little interactive communication at all between most members. Among the newsgroups that the researchers examined over a 130-day observation period, over half of the members did not contribute at all, and only a small number of the members contributed the majority of the messages (Cummings/Butler/Kraut 2002). So, while these kinds of Internet communities can (and often do) sustain ongoing interactions between individuals with fixed identities, a substantial number of interactions in these communities may actually consist of one-shot interactions between individuals who never communicate again.

5. Experimental Research on Trust and Implications for Research on the Internet

Current findings in experimental studies of trust formation are useful for understanding how networks of trust can form in computer-mediated systems such as the Internet. A series of risk taking behaviors is central to building a trust relation (Blau 1964; Holmes/Rempel 1989). In a typical trust building scenario, two people realize they can potentially gain from engaging in social exchange. Each person knows that she will gain from engaging in an exchange with a potential partner if that partner turns out to be trustworthy. However, there is not always a way to know if a given partner is actually trustworthy. On the Internet, for example, one does not know for certain if the person one is chatting with is who they say they are (i.e., male or female). Similarly, if two individuals engage in a trade over the Internet, there may be limited assurances that the other person will honor the terms of a trade.

Given such a risky situation, reputation becomes a useful and important tool for determining the trustworthiness of another person. This reputation may come in the form of either experience-based information (i.e., past experiences with this same individual), or third party-referenced information (i.e., reputation scores assigned by others who have interacted with the same individual) as noted above (Yamagishi/Matsuda 2003).

Risk-taking and One-Shot Interactions with Random Partners

If individuals have only one-shot interactions with one another and they are always new, random partners, trust cannot emerge. Even reputation information may not be possible to obtain in such a system because the interactions are always random. One solution is the use of formal or informal controls (Heckathorn 1993), yet such controls may not be possible either due to the costliness of maintaining them, or the impracticality of establishing them in open systems like those found on the Internet. For example, in the earlier example of an unsolicited email request, there are often no informal or formal mechanisms by which monitoring can be provided. This fact makes such interactions highly risky.

In a general study using simulations, Macy/Skvoretz (1998) show that cooperation and 'trust' can emerge, despite the lack of formal or informal control mechanisms, but only under certain conditions. Although the authors use the term 'trust' to refer to the willingness to participate in a risky activity (such as playing a PD game with a random partner), this is better referred to as *risk-taking behavior* (Hardin 2003) rather than *trust*. This research underscores the importance of distinguishing risk-taking from trusting behavior, especially in one-shot random partner situations where actual trust-building is not possible.

Macy/Skvoretz (1998) find that the emergence of cooperation among random partners in one-shot interactions ultimately depends on two structural conditions. First, individuals must have the option to *not* participate (or exit). Not only does an exit strategy better approximate the real world (since real people often have the ability to not interact at all if they so choose), but it also tends to encourage risk-taking if the option to exit is relatively costly.

The second condition they identify is the importance of embedded social networks. Macy/Skvoretz (1998) demonstrate that cooperation emerges locally among 'neighbors' in an exchange system. In this case, the term "neighbors" refers to the proximity of potential interaction partners in the system. Only after individuals cooperate in these local networks does cooperation disseminate through weak network ties. Thus, even if *trust* is not theoretically possible in this kind of one-shot random partner interaction situation, their research shows that the development of *cooperative behavior* is still possible under certain structural conditions.

Trust-building Among Fixed Partners in Repeated Interactions

In many situations there is often no regulatory agency to control behavior and little initial information about the reputations of possible exchange partners. Instead, individuals must determine the trustworthiness of their partners through experience. The process of building experience-based reputations is itself, risky—because each experience (i.e. interaction) offers a chance for one of the partners not to cooperate, or to betray the partner's trust. For example, in an online discussion among potential dating partners, each person takes a risk when they share increasingly personal information about themselves with the other person. Should they reveal their ages? Should they use their real names? Each decision is based on the results of previous interactions.

As mentioned above, our recent cross-cultural experiments on trust formation in the U.S. and Japan show that cooperation among individuals is significantly higher when individuals are given the opportunity to have repeated interactions with fixed partners instead of dealing with random partners (Cook et al. forthcoming). We also found that participants significantly entrusted more money when they were paired with fixed partners instead of random partners, and the cooperation rates were also higher when they were paired with fixed instead of random partners.

In the preceding example, individuals are simply not given the opportunity to exit the relationship. So, how does the situation change when individuals interact with a system of participants rather than a set of dyadic relationships? Yamagishi/Matsuda (2003) explore several questions about the trustworthiness of exchange partners in computer-mediated interactions by creating an Internet "auction" in a laboratory setting. The authors find that both experience-based *and* third-party reputations are successful in promoting trustworthy behavior in an exchange system[5]. However, these reputations are only effective so long as the identity of the individuals is *permanent*. The ability to change one's identity (a common feature in many forms of exchange on the Internet), undercuts the effectiveness of reputation systems by allowing individuals to cheat each other and then to start over with a new identity once their reputations become degraded.

[5] Yamagishi/Matsuda (2003) measured trustworthy behavior by examining the degree to which 'sellers' in their laboratory auction would sell goods at higher prices than they were worth. Since the true value of the goods was not known until after the transactions, these exchanges were analogous to many of the existing Internet auctions in which individuals buy and sell goods without inspecting the goods first.

In sum, *both* experience-based and third-party reputation information are essential for building trust relationships in computer-mediated interactions. While such reputation information is clearly important in the real world as well, it seems particularly crucial in online interactions that lack many of the cues that accompany real-world interactions.

Trust-building Among Random Partners in Repeated Interactions

What if individuals are either randomly matched or their identities change so that reputation information cannot be established as with repeated, fixed-partner exchange? When individuals are unable to use experience-based or third-party reputation information to secure their interactions, it becomes particularly difficult to ascertain the trustworthiness of their partners. And, it is especially difficult to facilitate the development of trust relationships through risk taking when there is no "shadow of the future" (Axelrod 1984). Trusting random partners, or those who can change their identity, begins to seem more like gambling than rational decision making. According to the encapsulated interest view of trust (Hardin 2002), such behavior would be defined as risk-taking, not trust.

We have mentioned that peer-to-peer exchange with digital goods represents one kind of Internet exchange situation where random partners can have repeated interactions. However, the apparent utopia of cooperative sharing in peer-to-peer systems may largely be artificially created by forcing users to share their own digital goods (Bricklin 2001). There are very real costs (i.e., time, network bandwidth, security) that might lead individuals to refrain from sharing digital goods if they had the choice. However, all of the most popular online file-swapping systems have this built-in feature: *users are not asked whether they want to share* their files. Instead, the files are automatically shared whenever someone downloads a digital good. As a result, the problem of coordinating cooperation among file-swappers is essentially eliminated.

If peer-to-peer systems did not artificially create this "cornucopia of the commons" (Bricklin 2001), would individuals still share digital goods? The answer to this question is not yet clear—but the long history of research on public goods and exchange does indicate that individuals do not, on the whole, contribute without some incentive for doing so. One possible motivation could be social psychological selective incentives that further reduce the already minor costs associated with sharing digital goods (Cheshire unpublished). Selective incentives have been very successful as an economic solution to free riding in the production of a public good because they offer additional motivations that make cooperation rational even when the initial conditions of the social dilemma make cooperation irrational (e.g., Olson 1965; Oliver 1980; Heckathorn 1996). Individuals may be encouraged to cooperate (i.e. share digital goods) if they receive social psychological rewards such as social approval and a sense of group solidarity when they contribute. In an experimental study of digital goods exchange, social psychological selective incentives are shown to have a significant influence on the likelihood of sharing, despite the costs associated with sharing and the lack of any specific reputation information about others (Cheshire unpublished).

In peer-to-peer Internet exchange and other types of situations in which random partners have repeated interactions, one of the key problems is the lack of reputation information. However, just because individuals cannot determine the reputation of a given partner does not mean that they cannot use other information or other tools to try to build trust. For example, we argue that the *signaling role* of trusting behavior is still important for building trust—even when individuals exchange with random partners. By acting in a trusting manner, one can signal her intention to cooperate. In a given system, this ability to signal may be particularly noticeable because it would be unexpected. For example, individuals who would probably prefer to cooperate with an exchange partner rather than to defect on her will often *defect anyway* because they expect that others are not willing to cooperate. These individuals may defect simply out of a fear of being exploited by others, not necessarily because of their own greed (cf., Pruitt/Kimmel 1977; Yamagishi/Sato 1986).

When individuals have random partners, there can also be a kind of "second-order" fear that others will be *similarly fearful* and thus they might defect for the same reason. It is this "second-order" fear that trusting behavior can lessen. For example, if two individuals meet on an Internet message board for book-swapping and decide to exchange books with each other, how do they know that they will each follow up on their agreement to mail the books (e.g., trusting behavior)? Furthermore, how can they be sure that the other person will return the book when they are finished (i.e., cooperative behavior)? When one of the individuals sends a book, it acts as a *signal* to the other person. By taking a risk and acting in a trusting manner, an individual signals to her partner that she expects cooperation, or, that she is not afraid that her partner will defect on the exchange. Thus, the *act of trusting* may help eliminate this second-order fear in her partner (see also Cook et al. forthcoming).

In our cross-cultural experiment on trust formation, we found that allowing individuals to signal their trustworthiness seems to have a positive effect on cooperation, but the effect is small and does not seem to last very long. In addition, we found that there is no cross-cultural difference in cooperation between the American and the Japanese participants when they participate in random-partner exchange. Thus, we do not find strong support for the argument that signaling is a mechanism for increasing cooperative behavior and trust in systems with random partners (Cook et al. forthcoming), but this is a fairly hostile environment for trust at the dyadic level. Further research varying the nature of the situation is needed to assess the importance of signaling at the population level.

6. Conclusion

Computer-mediated exchanges offer many new ways to look at social exchange and the development of cooperative behavior and trust. Many recent experimental studies on social exchange networks, uncertainty, commitment and trust are particularly useful for understanding the emerging forms of social interaction

that exist in computer-mediated situations like those found on the Internet and other mediated systems of trade. Since computer-mediated exchanges provide new possibilities for anonymity in different types of exchange situations, resolving the uncertainty and risk that are created by this anonymity is essential for creating ongoing, relatively secure exchange environments. As recent experimental research has shown, the development of trust networks (i.e. networks of those one views as trustworthy) can help resolve some of the difficulties involved in anonymous exchanges of different types of goods and services, but such networks have relatively high information requirements.

Across the various types of exchange networks, reputational information has consistently proven to be a key factor for establishing cooperation and facilitating the development of trust. However, other information such as the social context, community responsibility norms, or social psychological selective incentives can play an important role in trust-building and insuring reliability even when personal or third party reputation information is *not* available. As exchange on the Internet and other forms of computer-mediated interaction continue to evolve, ongoing experimental research in sociology and related disciplines offers insights into the structural and behavioral factors that influence the development of trust in these environments. We have offered a classification system that helps to link the types of exchange situations that are found in computer-mediated systems such as the Internet with the interaction situations that are the subject of much experimental research on cooperation and trust formation in the laboratory. Future research on the Internet and the many types of social and economic exchanges now being mediated through computerized connections will help inform not only theories of exchange, but also theories of the role of trust in the economy. In this case the real world has become the laboratory.

Bibliography

Abdul-Rahman, A./S. Hailes (2000), Supporting trust in virtual communities, in: *Proc. of 33rd Hawaii Int'l Conf on System Sciences, 6007*

Akerlof, G. A. (1970), The Market for Lemons: Quality Uncertainty and the Market Mechanism, in: *Quarterly Journal of Economics* 84, 488–500

Axelrod, R. (1984), *The Evolution of Cooperation*, New York

Beniger, J. (1987), Personalization of mass media and the growth of pseudo-community, in: *Communication Research* 14, 352–371

Berg, J./J. Dickhaut/K. A. McCabe (1995), Trust, Reciprocity, and Social History, in: *Games and Economic Behavior* 10, 122–42

Blau, P. M. (1964), *Exchange and Power in Social Life*, New York

Bricklin, D. (2001), The Cornucopia of the Commons, in: Andy Oram (ed.), *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, 59–66, Sebastopol, California

Cheshire, C. (2004), *A Sociological Analysis of Generalized Information Exchange*, Unpublished Ph.D. Dissertation

Cook, K. S./R. M. Cooper (2003), Experimental Studies of Cooperation, Trust, and Social Exchange, in: E. Ostrom/J. Walker (eds), *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, 209–244, New York

Cook, K. S./R. M. Emerson (1978), Power, Equity and Commitment in Exchange Networks, in: *American Sociological Review* 43, 721–739

—/R. Hardin (2001), Norms of Cooperativeness and Networks of Trust, in: M. Hechter/K. D. Opp (eds.), *Social Norms*, New York, 327–347

—/E. R. W. Rice/A. Gerbasi (2004), The Emergence of Trust Networks under Uncertainty: The Case of Transitional Economies–Insights from Social Psychological Research, in: S. Rose-Ackerman/B. Rothstein/J. Kornai (eds.), *Problems of Post Socialist Transition: Creating Social Trust*, 191–221, New York

—/T. Yamagishi/C. Cheshire/R. Cooper/M. Matsuda/R. Mashima (forthcoming), Trust Building via Risk Taking: A Cross-Societal Experiment, in: *Social Psychology Quarterly*

Cummings, J. N./B. S. Butler/R. Kraut (2002), The quality of online social relationships, in: *Communications of the ACM* 45, 7, 103–108

Curtis, P. (1991), *Mudding: Social Phenomena in Text-Based Virtual Reality*, Electronic manuscript, (FTP: parcftp.xerox.com), 1–21

Dasgupta, Partha (1988), Trust as a commodity, in: D. Gambetta *(ed.), Trust: Making and Breaking Cooperative Relations*, Oxford, 49–72

Donath, J. (1999), Identity and deception in virtual community, in: P. Kollock/M. Smith *(eds.)*, *Communities in Cyberspace,* 29–59, New York

Ekeh, P. (1974), *Social Exchange Theory: The Two Traditions*, Cambridge

Foa, U. G./E. B. Foa. (1974), *Societal Structures of the Mind*, Springfield

Gambetta, D./H. Hamill (forthcoming), *Streetwise: How Taxi Drivers Establish Customers' Trustworthiness*, New York

Grief, A. (1989), Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders, in: *Journal of Economic History* 49, 857–882

— (1993), Contract enforceability and economic institutions in early trade: the Maghribi Trader's coalitions, in: *American Economic Review* 83, 3, 525–548

Hardin, R. (1995), *One for All*, Princeton

— (2002), *Trust and Trustworthiness*, New York

Heckathorn, D. D. (1993), Collective Action and Group Heterogeneity: Voluntary Provision versus Selective Incentives, in: *American Sociological Review* 58,3, 329–350

— (1996), The Dynamics and Dilemmas of Collective Action, in: *American Sociological Review* 61, 250–277

Holmes, J. G./K. Rempel (1989), Trust in close relationships, in: C. Hendrick (ed.), *Close relationships*, 187–220, Thousand Oaks

Kollock, P. (1994), The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust, in: *American Journal of Sociology* 100, 2, 313–345

— (1999a), The Production of Trust in Online Markets, in: *Advances in Group Processes*, 99–123

— (1999b), The economies of online cooperation: Gifts and public goods in cyberspace, In: P. Kollock/M. Smith *(eds.)*, *Communities in cyberspace, 220–239*, London

Koster, R. (2003), Small Worlds: *Competitive and Cooperative Structures in Online Worlds*, conference talk, Game Developer's Conference, San Jose CA

Kreps, D. M (1990), Corporate Culture and Economic Theory, in J. E. Alt/K. A. Shepsle *(eds.), Perspectives on Positive Political Economy*, 90–143, Cambridge

Lessig, L. (1999), *Code and other laws of cyberspace*, New York

Macy, M. W./J. Skvoretz (1998), The Evolution of Trust and Cooperation Between

Strangers: A Computational Model, in: *American Sociological Review* 63, 638–660

Mori, K. (1996), The Effects of Trust and Communication on Cooperative Choice in a Two-Person Prisoner's Dilemma Game, in: *Japanese Journal of Experimental Social Psychology* 35, 3, 324–36

Oliver, P. (1980), Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations, in: *American Journal of Sociology* 85, 6, 1356–1375

Olson, M. (1965), *The logic of collective action; public goods and the theory of groups*, Cambridge

Pruitt, D. G./J. J. Kimmel (1977), Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future, in: M. R. Rosenzweig/L. W. Porter (eds.), *Annual review of psychology* 28, 363–392

Rheingold, H. (1993), *The Virtual Community*, New York

Rice, E. R. (2002), *The Effect of Social Uncertainty in Networks of Social Exchange*, Unpublished Ph.D. dissertation

Sato, K. (1988), Trust and Group Size in Social Dilemmas, in: *Japanese Psychological Research* 30, 2, 88–93

Smith, L./P. Sørensen (2000), Pathological outcomes of observational learning, in: *Econometrica* 68, 371–398

Snijders, C. (1996), *Trust and Commitments*. Utrecht: Interuniversity Center for Social Science Theory and Methodology

Sproull, L./S. Kiesler (1986), Reducing social context cues: electronic mail in organizational communication, in: *Management Science* 32, 11, 1492–1512

Stefik, M. (ed.) (1999), *The Internet Edge: Social, Technical, and Legal Challenges for a Networked World*, Cambridge, Massachusetts

Takahashi, N. (2000), The Emergence of Generalized Exchange, in: *American Journal of Sociology* 105, 1105–1134

Wellman, B./G. Milena (1999), Virtual communities as communities: Net surfers don't ride alone, in: Kollock, P./M. Smith (eds.), *Communities in Cyberspace*, 167–194, New York

Whitmeyer, J. M. (2000), Power over Groups through Effective Monitoring and Sanctioning, conference presentation: *American Sociological Association*

Xiong, L./L. Liu (2003), A Reputation-Based Trust Model for Peer-to-Peer eCommerce Communities, Unpublished Manuscript, 2003 IEEE *Conference on E-Commerce (CEC'03)*, Newport Beach, USA

Yamagishi, T./M. Matsuda (2003), The role of reputation in open and closed societies: An experimental study of online trading. *Center for the Study of Cultural and Ecological Foundations of Mind, Working Paper Series*, 8

Yamagishi, T./K.S. Cook (1993), Generalized Exchange and Social Dilemmas, in: *Social Psychology Quarterly* 56, 235–248

—/Kaori Sato (1986), Motivational bases of the public goods problem, in: *Journal of Personality and Social Psychology* 50, 67–73

—/R. Kakiuchi (2000), It takes venturing into a tiger's cave to steal a baby tiger: Experiments on the development of trust relationships, in: W. Raub/J. Weesie (eds.), *The Management of Durable Relations*, 121–123, Amsterdam