

**RATIONAL CHOICE RESEARCH ON SOCIAL DILEMMAS:
EMBEDDEDNESS EFFECTS ON TRUST[†]**

Vincent Buskens

Department of Sociology / ICS
Utrecht University
Heidelberglaan 2
3584 CS Utrecht, Netherlands
v.buskens@uu.nl

Werner Raub

Department of Sociology / ICS
Utrecht University
Heidelberglaan 2
3584 CS Utrecht, Netherlands
w.raub@uu.nl

Version April 18, 2008

Word count: about 24,000 (including references and footnotes)

Prepared for:

Rafael Wittek, Tom A.B. Snijders & Victor Nee (eds.)
Handbook of Rational Choice Social Research,
New York: Russell Sage 2008.

[†] Stimulating comments of and discussions with Jeroen Weesie and other members of our Utrecht group “Cooperation in Social and Economic Relations” are gratefully acknowledged. We also acknowledge helpful comments from participants of the Russell Sage Foundation “Rational Choice Social Research Workshop” and specifically from our discussant, Simon Gächter. Financial support for Buskens was provided by the Royal Netherlands Academy of Arts and Sciences (KNAW) for the project “Third-Party Effects in Cooperation Problems” and by Utrecht University for the High Potential-program “Dynamics of Cooperation, Networks, and Institutions.” Financial support for Raub was provided by the Netherlands Organization for Scientific Research under grants S 96-168 and PGS 50-370 for the PIONIER-program “The Management of Matches” and under grant 400-05-089 for the project “Commitments and Reciprocity.” The order of authorship is alphabetical. This chapter will be accompanied by an Internet based appendix with suggestions for further reading and other supplementary material at www.dyconi.nl/socialdilemmas.htm.

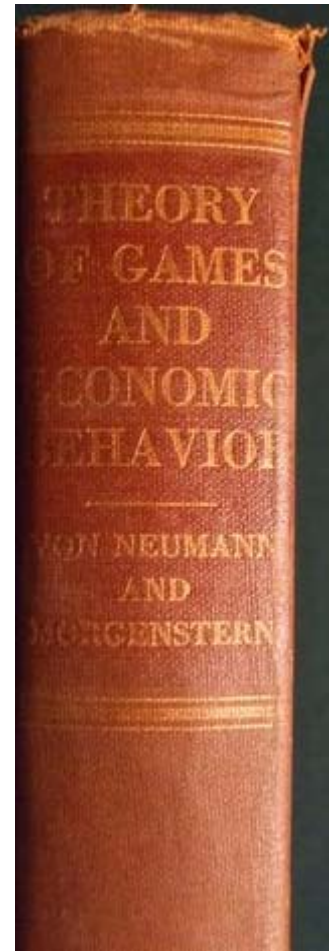
Rational Choice Research on Social Dilemmas: Embeddedness Effects on Trust

Vincent Buskens and Werner Raub

INTRODUCTION

Social Dilemmas by Example: Trust in Economic Exchange

Consider economic exchange through the Internet. July 18, 2007 was the end date to purchase a copy of the first edition of *Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern at eBay from the seller “bibliomonster” for US-\$ 1,900.00. The item had a fixed price listing (using eBay’s “Buy It Now” option) and could only be purchased without bidding in an auction. Assume that the seller did own the copy, that the description of the copy was accurate (“Bound in original publishers red cloth a bit rubbed at head of spine. Black (ink?) mark on top board. Minor shelf wear, else very good. Internally, clean and free of ink, marginalia and soiling. No dogeared pages or tears. Includes the often missing corrigenda leaf. A nice, collectable copy.”), and that the accompanying photos were not misleading. Assume that a buyer existed who would have preferred purchasing the copy as described for the price mentioned. The buyer had to pay the price before the seller would ship the book. Thus, the buyer had to trust the seller that the copy would indeed be shipped. If the buyer decided not to pay, there would be no transaction. If the buyer decided to pay, the seller had to decide whether or not to ship the copy.¹



As a benchmark scenario, imagine an “isolated encounter,” i.e., a one-shot transaction in the sense that buyer and seller of the book have never done business with each other before, do not expect to do business with each other in the future, and that verifying the seller’s identity is prohibitively costly for the buyer. For the benchmark, imagine further that eBay would not maintain its feedback forum that allows buyers to evaluate sellers, with evaluations being publicly available and easy to access. As a second and less artificial scenario for buying antiquarian books, consider that websites of antiquarian booksellers typically offer indications of their identity, such as information on the physical location of their shop. Imagine, too, that the buyer has purchased antiquarian books from the seller in the past, and the seller expects that the buyer may purchase such books also in the future. Finally, consider a third scenario that takes a core feature of the eBay platform into account, namely eBay’s feedback forum that provides information on the seller from other buyers. On July 16, 2007, the seller of the first edition of *Theory of Games and Economic Behavior* had positive feedback from 386 other eBay members and negative feedback from 5 members, resulting in an eBay feedback score

¹ We decided not to purchase the copy ourselves and the item was not sold.

of 381, with 98.7% positive feedback. Moreover, the seller was a “PowerSeller” (“PowerSellers are eBay top sellers who sustain a consistent high volume of monthly sales and a high level of total Feedback – with 98% or better positive rating by other eBay users”).

Trust as a Social Dilemma

Our example of the purchase of a first edition of *Theory of Games and Economic Behavior* represents a trust problem between the buyer and the seller in Coleman’s (1990, 97–99) sense, with the buyer in the role of the trustor and the seller in the role of the trustee. Coleman emphasizes four points that characterize a trust problem:²

- (1) Placing trust by the trustor allows the trustee to honor or abuse trust, while this alternative is not available for the trustee without placement of trust. In the example, if the buyer decides to buy, the seller can ship the copy of the first edition or can abstain from shipping. In addition, while Coleman does not mention this explicitly, it is important that the trustee has not only an opportunity but also an incentive to abuse trust. For example, the seller could change his or her virtual identity and offer the book once again through the Internet.
- (2) Compared to the situation with no trust placed, the trustor is better off if trust is placed and honored but is worse off if trust is abused. The buyer prefers to purchase the first edition to, say, owning only the 1980 Princeton paperback printing, while owning the paperback version only is most likely still preferred by the buyer to paying US-\$ 1,900.00 without receiving the first edition. Again in addition to Coleman’s characterization, the trustee is better off if trust is honored than if no trust is placed. Selling the book for US-\$ 1,900.00 is profitable for the seller.
- (3) There is no “real commitment” (Coleman 1990, 98) of the trustee to honor trust. Thus, the trustor voluntarily places resources in the hands of the trustee. In the benchmark scenario, since the buyer cannot verify the identity of the seller, she is not able to enforce shipment of the copy after payment.
- (4) There is a time-lag between placement of trust by the trustor and the action of the trustee. The buyer first pays the price and, subsequently, the seller decides on whether or not to ship the book.

In the benchmark scenario resembling a one-shot transaction, it seems intuitive that incentive-guided and goal-directed behavior of trustor and trustee implies that the trustee would indeed abuse trust, if trust is placed. Assuming that the trustor anticipates this, she does not place trust in the first place. If trust is not placed, however, both trustor and trustee are worse off than when trust is placed and honored. Technically speaking, the no trust-outcome is Pareto-suboptimal. As Rapoport (1974) aptly put it, individual rationality in the sense of incentive-guided and goal-directed behavior can lead to collective irrationality in the sense of Pareto-suboptimality. Such a “conflict” between individual and collective rationality is the core feature of a social dilemma and trust relations are a paradigmatic example of a social dilemma involving two actors. While “social dilemma” is a label commonly used in social psychology and also sociology, such a situation is often referred to as a “problem of collective action” or the “tragedy of the commons” in political science and as a “public goods problem” in economics (see Ledyard 1995, 122). Below, we explain why we focus on trust problems in this chapter.

² This characterization of trust is close to Hardin’s (2001, 2002, chap. 1) conceptualization of “trust as encapsulated interest.”

Social dilemmas are intimately related to the problem of order in Parsons' (1937) sense. After all, in Hobbes' ([1651] 1991, chap. 13) "naturall condition of mankind," actors are interdependent in a world of scarcity, while binding and externally enforced contracts are unfeasible. They may thus end up in the "warre of every man against every man." In that situation, the life of man is "solitary, poore, nasty, brutish, and short" and everybody is worse off compared to a peaceful situation. This is a social dilemma among many actors. Parsons (1937, 89–94) posed the challenge to specify conditions such that individually rational actors solve the problem of order. He thus (1937, 91) referred to the problem of order as "the most fundamental empirical difficulty of utilitarian thought." In his meanwhile classic early contribution to rational choice social research, Coleman (1964, 166–67) clearly realized the challenge and formulated it even more radically than Parsons: "Hobbes took as problematic what most contemporary sociologists take as given: that a society can exist at all, despite the fact that individuals are born into it wholly self-concerned, and in fact remain largely self-concerned throughout their existence. Instead, sociologists have characteristically taken as their startingpoint a social system in which norms exist, and individuals are largely governed by those norms. Such a strategy views norms as the governors of social behavior, and thus neatly bypasses the difficult problem that Hobbes posed [...] I will proceed in precisely the opposite fashion [...] I will make an opposite error, but one which may prove more fruitful [...] I will start with an image of man as wholly free: unsocialized, entirely self-interested, not constrained by norms of a system, but only rationally calculating to further his own self interest."

While it is part of the sociological folklore that Parsons' challenge focuses on how rational choice social research can cope with social dilemmas, it is less well appreciated that Durkheim put forward a similar argument in his analysis of the division of labor in society ([1893] 1973, book I, chap. 7) that nicely relates to the antiquarian book example. Durkheim's point is that economic transactions often deviate from what is conventionally assumed in standard neo-classical models of spot exchange on perfect markets. Specifically, Durkheim highlighted the limits of "contractual governance" of economic transactions. Durkheim argued that the governance of transactions exclusively via bilateral contracts requires that the present and future rights and obligations of the partners involved in the transaction are specified explicitly for all circumstances and contingencies that might arise during and after the transaction. Anticipating much of the modern economic and game-theoretic literature on incomplete and implicit contracts, Durkheim pointed out that such purely contractual governance of economic transactions is problematic: Typically, many unforeseen or unforeseeable contingencies could or actually do arise during or after a transaction. Negotiating a contract explicitly covering all these contingencies would be unfeasible or at least prohibitively costly. Likewise, renegotiations in the case that contingencies arise are also costly (for similar arguments on the limits contractual governance, see Weber [1921] 1976, 409 in his sociology of law). Such renegotiations characteristically offer incentives for opportunistic behavior since an unexpected contingency will often strengthen the bargaining position of one partner while weakening the position of the other. Hence, Durkheim argues that mutually beneficial economic exchange presupposes the solution of a trust problem and thus involves a social dilemma.

Focus of the Chapter

Parsons and Durkheim quite convincingly show that social dilemmas are a strategic research site in Merton's (1973) sense for rational choice social research. Game-theoretic models have emerged as a major tool for the analysis of social dilemmas in rational choice

social research. This is not accidental. Interdependence between actors is a core feature of a social dilemma. For example, the behavior of the trustor has effects for the trustee and vice versa. Game theory is the branch of rational choice theory that models interdependent situations, providing concepts, assumptions, and theorems that allow to specify how rational actors behave in such situations. The theory assumes that actors behave as if they try to realize their preferences in decision situations with restrictions, taking their interdependencies as well as rational behavior of the other actors into account (e.g., Harsanyi 1975, 89–117). It is therefore natural that applications of game-theoretic models figure prominently in rational choice social research on social dilemmas. Moreover, one should observe that interdependencies between actors and actors taking their interdependencies into account are likewise the core of Weber's (1947, 88, emphasis added) famous definition of social action: "Sociology [...] is a science which attempts the interpretive understanding of social action in order thereby to arrive at a causal explanation of its course and effects [...] Action is social in so far as [...] *it takes account of the behaviour of others and is thereby oriented in its course.*" This is a reason why social dilemmas are a strategic research site not only for rational choice social research but for sociology in general and why game theory is an important tool for rational choice social research and, more broadly, for sociological analyses in the spirit of Weber.

Reviews of social dilemma research are readily available that highlight how social psychological theory and other approaches with a firm basis in methodological individualism can be used in this field that differ from rational choice assumptions (e.g., Kollock 1998). Somewhat surprisingly, though, there is no systematic review of applications of game theory in this field with a focus on how sociologically informed hypotheses can be derived from these models and how these hypotheses fare in empirical research. The present review contributes to filling this gap. We do so by reconstructing research strategies that are often employed in applications of game theory for the analysis of social dilemmas as well as reconstructing core assumptions and implications, including testable hypotheses. The chapter is analytical in nature, trying to structure the field, rather than providing a general overview of and comparison with alternative theories. Suggesting that theoretically and empirically informed middle range theory on social dilemmas has emerged from applications of game theory in this field, empirical insights generated by theoretical models are a core topic of the chapter. In the spirit of Goldthorpe's (2000, chap. 5) plea for an alliance between rational action theory (RAT) and the quantitative analysis of data (QAD), we explore how research in this field can contribute to narrowing the gap between rational choice models and empirical research (Green and Shapiro 1994). The review emphasizes that relevant empirical research in the field includes survey designs, quasi-experimental designs, and more qualitative case studies in addition to experimental designs. Such a "multi-method" perspective conceives QAD broadly and is particularly appropriate when it leads to testing similar hypotheses with complementary empirical designs, thus providing an indication of the robustness of empirical findings.

This Handbook (see Introduction) focuses on two strategies through which rational choice theory has been extended beyond the highly stylized assumptions of neo-classical economics, namely, atomized interaction on perfect markets of rational and selfish actors with full information. One strategy involves making the assumptions on the actors more complex by relaxing the rationality assumption or the selfishness assumption (see Gächter's chapter in this Handbook on how this strategy can be usefully employed for improving on standard models and applications of game theory in rational choice social research). The other strategy aims at using more complex and more appropriate assumptions on the social context by replacing the assumption of atomized interactions on perfect markets. This chapter highlights the second

strategy.³ It does so by combining strong assumptions on individual rationality with assumptions on the “embeddedness” of action in ongoing relations, networks of relations, and institutions, showing that embeddedness crucially affects behavior of rational actors in social dilemmas. This is not only in line with Coleman’s (1987) heuristic advice to combine robust assumptions on rational behavior with more complex assumptions on social structure. Also Granovetter (1985) advocated precisely such a combination of assumptions in his often cited programmatic sketch. Granovetter’s criticism of the shortcomings of the neo-classical model of perfect markets with atomized actors has often been taken to imply that one had better abandon rational choice models in favor of more “realistic,” socially inspired models of man. It has been widely overlooked, though, that Granovetter sharply opposes “psychological revisionism” characterizing it as “an attempt to reform economic theory by abandoning an absolute assumption of rational decision making” (1985, 505). Rather, he suggests to maintain the rationality assumption: “[W]hile the assumption of rational action must always be problematic, it is a good working hypothesis that should not easily be abandoned. What looks to the analyst like nonrational behavior may be quite sensible when situational constraints, especially those of embeddedness are fully appreciated” (1985, 506). He argues that investments in tracing the effects of embeddedness are more promising than investments in the modification of the rationality assumption: “My claim is that however naive that psychology [of rational choice] may be, this is not where the main difficulty lies—it is rather in the neglect of social structure” (1985, 628). This chapter explores the potential of such an approach for the analysis of social dilemmas.

We use trust problems as a paradigmatic example of social dilemmas, sometimes indicating generalizations of results for other types of social dilemmas. Trust problems involve two actors. We thus largely neglect social dilemmas involving many actors (these are covered in Heckathorn’s chapter in this Handbook). Our chapter first sketches core concepts and assumptions of game theory. We illustrate how game-theoretic tools can be used for modeling trust problems and other social dilemmas and sketch the logic of deriving testable hypotheses from game-theoretic models. We then turn to theory and hypotheses on how social structure, i.e., embeddedness of a trust problem or, more generally, embeddedness of a social dilemma affects behavior in such situations. The review of empirical research takes stock of evidence for and against hypotheses on embeddedness effects, with an emphasis on results obtained from complementary research designs and an emphasis on applications to the Internet economy and other economic exchange that resembles a social dilemma. Some directions for future research are also suggested.

GAME-THEORETIC TOOLS

Game theory provides tools for the analysis of situations with interdependence of two or more actors: choices of an actor affect the other actor(s) and vice versa. We sketch concepts and assumptions of game theory used in this section.⁴ We proceed informally and rely on examples. Our aim is to foster intuition rather than technical precision with respect to terminology or proofs of theorems.⁵

³ See Ostrom (2003) for first steps towards a theoretical framework that combines both strategies.

⁴ We focus exclusively on non-cooperative games. These are games in which actors cannot make binding agreements or binding unilateral commitments that are not explicitly modeled as moves in the extensive form of the game (see below). As will become transparent, this does not exclude that actors behave cooperatively in a non-cooperative game. We show that models of non-cooperative games are useful, among other things, because they allow for an analysis of conditions for cooperation.

⁵ For a textbook accessible for readers with modest training in formal theoretical model building and no prior exposure to game theory, see, e.g., Rasmusen (2007).

Some Core Concepts of Game Theory

Consider the standard *Trust Game* (Camerer and Weigelt 1988; Dasgupta 1988; Kreps 1990a; Snijders 1996, chaps. 1–4; Buskens 2002, chaps. 1–3) that models trust problems as outlined above. The game (see Figure 1) involves two actors, the trustor (actor 1) and the trustee (actor 2). Games with more than two actors can be likewise analyzed.⁶ The game starts with a move of the trustor. She can choose between placing or not placing trust. If trust is not placed, the interaction ends and the trustor receives payoff P_1 , while the trustee receives payoff P_2 . If trust is placed, the trustee chooses between honoring and abusing trust. If he honors trust, the payoffs for trustor and trustee are $R_i > P_i$, $i = 1, 2$. If trust is abused, the payoff for the trustor is $S_1 < P_1$, while the trustee receives $T_2 > R_2$.

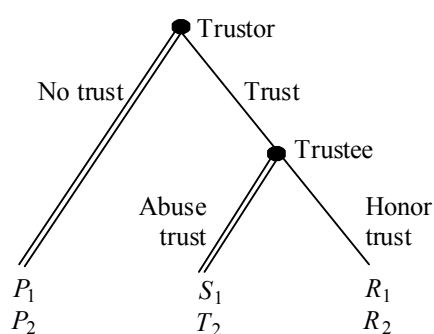


FIGURE 1. The Trust Game ($S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$)

The Trust Game models the benchmark scenario of a one-shot transaction between a buyer and a seller of an antiquarian book. The buyer is the trustor and chooses between placing trust by paying the price for the book and not placing trust by not buying. The seller is the trustee who can honor trust by shipping the book or abuse trust by not shipping. Of course, one could imagine variants of the game. For example, rather than not shipping the book, abusing trust could mean that the seller ships a low quality version of the first edition or even a reprint.

A slightly more complex model of a trust problem is the *Investment Game* (Berg, Dickhaut, and McCabe 1995; Ortmann, Fitzgerald, and Boeing 2000; Barrera 2005). Again, the game is played by two actors. However, while the actors make binary choices in the Trust Game, the trustor now chooses the degree to which she trusts the trustee in the Investment Game. The trustee now chooses the degree to which he honors trust. More precisely, the trustor has an endowment E_1 and chooses an amount M_1 to send to the trustee ($0 \leq M_1 \leq E_1$). This “investment” M_1 is then multiplied by $m > 1$ and the trustee receives mM_1 . The parameter m can be seen as indicating the trustee’s returns due to the trustor’s investment. Subsequently, the trustee chooses an amount K_2 he returns to the trustor, with $0 \leq K_2 \leq mM_1$. Afterwards, the game ends with the trustor receiving $V_1 = E_1 - M_1 + K_2$ and the trustee receiving $V_2 = mM_1 - K_2$. While M_1 indicates how much the trustor trusts the trustee, K_2

⁶ Actors do not need to be natural persons in applications of game theory. Rather, and contradicting an extreme version of methodological individualism, applications quite often involve the simplifying assumption that “corporate actors” in Coleman’s (1990) sense, such as firms or states, are conceived as actors.

indicates how trustworthy the trustee's is.⁷ Both the Trust Game and the Investment Game represent the time lag between placing trust and the trustee's response. In both games, trust is risky because the trustor regrets her trustfulness if the trustee turns out not to be trustworthy.⁸ The games thus model risks for the trustor in the sense of "opportunistic" or "strategic" behavior of the trustee who has an incentive for abusing trust. Conversely, these games do not model trust in the sense of "confidence" of the trustor in competencies of the trustee (Barber 1983). In the example, we focus on the risk for the buyer that the seller does not ship the book, while we neglect the possibility that he intends to honor trust and ships the book but that the book arrives damaged because of expensive but unprofessional packaging.

Figure 1 provides the *extensive form* of the Trust Game. The extensive form is a *tree-like* representation. At each *node*, the *actor (player)* is indicated who has to make a move. A *move* of an actor is an action of that actor at some node. At the endnodes of the tree, the actors' *payoffs* are indicated. We interpret these payoffs as cardinal utilities. Game theory *as such* does not include assumptions on whether an actor's utility depends exclusively on the actor's own material and possibly monetary outcomes. Neither does game theory *as such* involve assumptions on, for example, an actor's risk preferences. Thus, in principle, additional assumptions that an actor's utility (also) depends on the outcomes of an interaction for the other actor, on the fairness of outcomes, on "social orientations," etc. are consistent with using a game-theoretic framework. It is now also transparent in what sense trustor and trustee are interdependent. For example, in the Trust Game, the trustee's behavior determines whether the trustor obtains payoff R_1 or S_1 after trust has been placed. On the other hand, the trustor's behavior determines whether the trustee obtains P_2 or, respectively, can attain T_2 or R_2 . We assume throughout that the extensive form of the game is *common knowledge* for the actors involved: it is known to each actor, each actor knows that it is known to each actor, and so forth.⁹

A core concept of game theory is an actor's *strategy*. This is a rule specifying the actor's behavior at each node where the actor has to move (this is a rough characterization that will be refined below). In the Trust Game, there is only one such decision node for each actor. This could create the impression that a strategy resembles a move. In general, however, a strategy is much more "complex," namely, a plan of how to behave during the game that specifies the actor's behavior for all circumstances that could emerge. This can be seen already in the Investment Game. A strategy for the trustee specifies his move at *each* of the nodes where he can end up, depending on the prior move of the trustor.

Game-Theoretic Assumptions

Up to now, we have sketched terminology. However, a theory is not a system of concepts but a system of propositions. Thus, we now turn to assumptions used in game theory. Game theory assumes rational behavior of the actors (see Gintis 2007 on how to interpret rationality assumptions and for a recent and nuanced discussion of objections to rationality assumptions; we try to keep the exposition brief and thus neglect, e.g., "evolutionary" underpinning of rationality assumptions in the sense of Gintis; see also Fudenberg and Levine

⁷ If the amount M_1 sent by the trustor is "small," a "small" amount K_2 returned by the trustee could also be interpreted as a punishment the trustee inflicts on the trustor for not trusting the trustee.

⁸ Yet another game with a similar strategic structure is the gift-exchange game (Fehr, Kirchsteiger, and Riedl 1993) that has been used for experimental research on wage setting by employers and subsequent effort levels by employees.

⁹ The common knowledge assumption is neither unproblematic nor always necessary (see, e.g., Gintis 2007). It is used here to keep the exposition brief.

1998). Each actor is assumed to use a strategy that maximizes the actor's expected utility given the opportunities and restrictions as represented in the extensive form. How to specify such an assumption, though, is more problematic in situations with interdependence between actors than in situations without interdependence (e.g., lotteries). After all, the consequences for an actor of choosing a certain strategy also depend on the strategies chosen by other actors and vice versa. Thus, an actor has to anticipate on other actors' behavior and vice versa. The problem emerges that actors have to form rational expectations with respect to the behavior of other actors, whose behavior likewise depends on such expectations.

This shows that specifying the *solution* of a game, i.e., the combination of strategies chosen by rational actors, is less than trivial. Nash's (1951) equilibrium concept is a key contribution towards tackling this task. A *Nash equilibrium* is a combination of strategies, one for each actor, such that each actor's strategy is a best reply against the strategies of the other actors in the sense that it maximizes the actor's expected payoff against those other strategies. Thus, no actor has an incentive to deviate unilaterally from his or her own equilibrium strategy, given that the other actors use their equilibrium strategies. One easily verifies that the Trust Game has a unique equilibrium such that the trustee would abuse trust, while the trustor does not place trust (in a sense, anticipating on the trustee's abuse of trust if trust would be placed). This is the game-theoretic underpinning of the intuition that the seller would not ship the first edition in our benchmark scenario and that the buyer would rather not purchase the copy in the first place. In Figure 1, double lines indicate the actors' moves in the equilibrium, while the payoffs associated with the equilibrium are encircled in the respective cell in Figure 2. Simple, though strong, assumptions on properties of the solution of a game imply that the solution has to be a Nash equilibrium. Namely, assume (1) that a game has a unique solution (assume, thus, that the concept of rational behavior is well defined), (2) that all actors behave as if they anticipate the solution, and (3) that all actors are rational. It then follows that the solution has to be a Nash equilibrium. Thus, Nash equilibrium behavior is the basic game-theoretic specification of individual rationality à la Rapoport.

One of Nash's major achievements that secured him the 1994 Nobel Prize in economics for his work in game theory (together with Selten and Harsanyi whose major contributions are sketched below) is the proof of the existence of a Nash equilibrium for a broad class of games, namely, finite games. These are games with a finite number of actors and a finite number of (pure) strategies for each actor.¹⁰ However, many games have more than one Nash equilibrium. Hence, being a Nash equilibrium is a necessary but not a sufficient condition for the solution of a game. This *equilibrium selection problem* is commonly seen as a, if not *the*, major fundamental problem of game theory (see Kreps 1990b, chap. 5 for an accessible discussion).

The equilibrium selection problem emerges already in the Investment Game. One easily verifies that the game has many Nash equilibria, namely, all strategy combinations such that the trustor sends nothing (she chooses $M_1 = 0$), while the trustee's strategy is such that for all $M_1 > 0$ he chooses some $K_2 \leq M_1$. All these equilibria do imply that nothing is sent in equilibrium but there are many strategies of the trustee that are best replies against the equilibrium strategy of the trustor to send nothing. The reason is that the trustee's best reply strategies against the trustor's strategy to send nothing differ with respect to what they require the trustee to do off the equilibrium path, i.e., should the trustor send more than nothing.

¹⁰ In this chapter, "strategy" always refers to a *pure strategy* in the technical game-theoretical sense. An actor plays a *mixed strategy* when choosing randomly between pure strategies. For example, the trustee uses a mixed strategy in the Trust Game when he would honor trust with a probability p after trust has been placed and would abuse trust with the complementary probability $1 - p$. Nash proved that if one allows for mixed strategies, each finite game has at least one equilibrium.

Conversely, as long as the trustee uses a strategy that assures that he always returns less than the trustor has sent, the trustor's best reply is to send nothing.¹¹

The equilibrium selection problem indicates that the solution of a game has to fulfill additional criteria beyond being a Nash equilibrium. To this end, various *refinements* of the Nash equilibrium concept have been developed. The most important refinement is Selten's (1965) *subgame-perfect equilibrium*. For an intuitive understanding of subgame perfection, consider a part of a game tree that can itself be considered as a game tree and, thus, represents a subgame of the original game. For example, in the Trust Game, the tree that starts with the node where the trustee chooses between honoring and abusing trust represents a subgame of the Trust Game. In the Investment Game, subgames start at each node where the trustee decides on how much to return to the trustor. A subgame-perfect equilibrium is a strategy combination that is a Nash equilibrium for the game and *also* for each subgame. Selten proved the existence of subgame-perfect equilibria for finite games.

The intuitively attractive property of subgame-perfect equilibria is that they comprise credible promises and, in particular, credible threats. Obviously, in the Trust Game, the equilibrium such that the trustee would abuse trust while the trustor does not place trust is the unique subgame-perfect equilibrium. The Investment Game likewise has a unique subgame-perfect equilibrium, namely, the trustee never returns anything ($K_2 = 0$ for all nodes where the trustee moves), while the trustor sends nothing. Note that all equilibria of the Investment Game that are not subgame perfect comprise an (implicit) promise of the trustee that is not credible. Namely, the strategy of the trustee in a Nash equilibrium of the Investment Game that is not subgame perfect implies that he returns some $K_2 > 0$ at least at some node, while this would not maximize the trustee's payoff at that node.¹² From a game-theoretic perspective, subgame-perfect equilibrium behavior is a natural further specification of individual rationality.

Assuming rational behavior of the actors, credible threats typically need not be implemented, precisely because they are credible. The concept of a subgame-perfect equilibrium and the intuitive justification of subgame perfection in the sense of credible threats and promises thus reveals that whether or not a rational actor uses a certain strategy can depend significantly on what the strategy requires to do in situations that will actually never emerge, given the solution. Observe that double lines in Figure 1 indicate behavior induced by the subgame-perfect equilibrium. A double line that runs from the starting node of a game to an endnode indicates the equilibrium path of play, i.e., the behavior that results when the actors implement their subgame-perfect equilibrium strategies. The Investment Game is a game with many Nash equilibria and a unique subgame-perfect equilibrium. In such a case, the subgame-perfect equilibrium can be considered as the solution of the game. However, we will see that games can have more than one – and possibly many – subgame-

¹¹ Allowing for mixed strategies, the Trust Game likewise has Nash equilibria in addition to the equilibrium in pure strategies of abusing and not placing trust. The set of Nash equilibria is the set of all strategy combinations such that the trustor does not place trust while the trustee would honor trust with probability $p \leq (P_1 - S_1)/(R_1 - S_1)$. For the Investment Game, one likewise easily constructs additional Nash equilibria that involve mixed strategies for the trustee. Although multiple equilibria exist in the Trust Game and in the Investment Game, they only differ in the specification of behavior at nodes that will not be reached in equilibrium: behavior is the same in all equilibria of the Trust Game and the Investment Game, i.e., no trust is placed. In other games, though, different equilibria in general do differ with respect to the behavior of the actors that is induced by the respective equilibrium strategies.

¹² Similarly, in the Trust Game, the equilibria with mixed strategies of the trustee (see previous note) are not subgame perfect.

perfect equilibria. In general, therefore, subgame perfection is again at best a necessary but not a sufficient property for solution strategy combinations.¹³

An Extension: Incomplete Information

The models considered so far comprise the assumption that actors are completely informed on all features of the game. This assumption is often problematic for social interactions. For example, a buyer at an eBay auction is less than completely informed on the behavioral alternatives or incentives of the seller. Does the seller indeed own the product that is auctioned? Could the seller ship a low-quality variant of the product? Alternatively, the trustor might be incompletely informed on the trustee's incentives. More specifically, the trustee might derive more utility from honoring trust than from abusing trust due to the internalization of norms and values that imply "internal sanctions" when trust is abused. Under such *incomplete information*, the trustor can no longer be sure about the behavior of the trustee after placement of trust. Game theory includes tools for modeling a *Trust Game with incomplete information* (see Rasmusen 2007, chap. 2 for an introduction to the theory of games with incomplete information and Bacharach and Gambetta 2001 for arguments why the Trust Game with incomplete information may be a more adequate model of trust problems than the Trust Game with complete information in Figure 1). The extensive form of the game is shown in Figure 3. The game starts with a random move of Nature that determines the *type* of the trustee a trustor encounters. With probability π , the trustee's payoff from abusing trust is $T_2^* < R_2$ (e.g., this trustee has internalized norms and values of trustworthiness), while with probability $1 - \pi$ the trustee's payoff from abusing trust is $T_2 > R_2$. The trustee knows his own incentives, while the trustor is only informed on the probability π and cannot directly observe the outcome of the random move of Nature. In Figure 3, this is indicated by including the two nodes for the move of the trustor in one information set represented by the line around those nodes. The trustor does not know at which of these two nodes she has to move. We have to refine our characterization of what a strategy is by saying that a strategy specifies how an actor moves at each information set where the actor has to move. This shows how incomplete information can be modeled and how random elements ("external contingencies") can be included in game-theoretic models.

It is straightforward to identify the subgame-perfect equilibrium of the Trust Game with incomplete information. A trustee with internalized norms and values of trustworthiness honors trust, while the other type of trustee abuses trust. If the trustor does not place trust, she receives P_1 , while her expected payoff from placing trust is $\pi R_1 + (1 - \pi)S_1$. The trustor's unique equilibrium strategy is thus not to place trust if $\pi < (P_1 - S_1)/(R_1 - S_1)$. Conversely, placing trust is the unique equilibrium strategy if $\pi > (P_1 - S_1)/(R_1 - S_1)$. If $\pi = (P_1 - S_1)/(R_1 - S_1)$, the game has multiple subgame-perfect equilibria since both trustor's strategies are best replies.¹⁴ Note that $(P_1 - S_1)/(R_1 - S_1)$ is a convenient measure of the risk the trustor incurs

¹³ The "second-order free-rider problem" (e.g., Coleman 1990, chap. 11) refers to the problem of providing sufficient incentives for the implementation of sanctions. Subgame-perfection of an equilibrium assures the solution of this problem.

¹⁴ Coleman's (1990, chap. 5) well-known condition for placing trust is equivalent to $\pi > (P_1 - S_1)/(R_1 - S_1)$. Thus, it is not necessary to introduce this condition as an assumption. Rather it can be derived from a game-theoretic model as an implication.

when placing trust. Our example for a game with incomplete information is simple. Harsanyi (1967/68) has shown how games with incomplete information can be analyzed in general.¹⁵

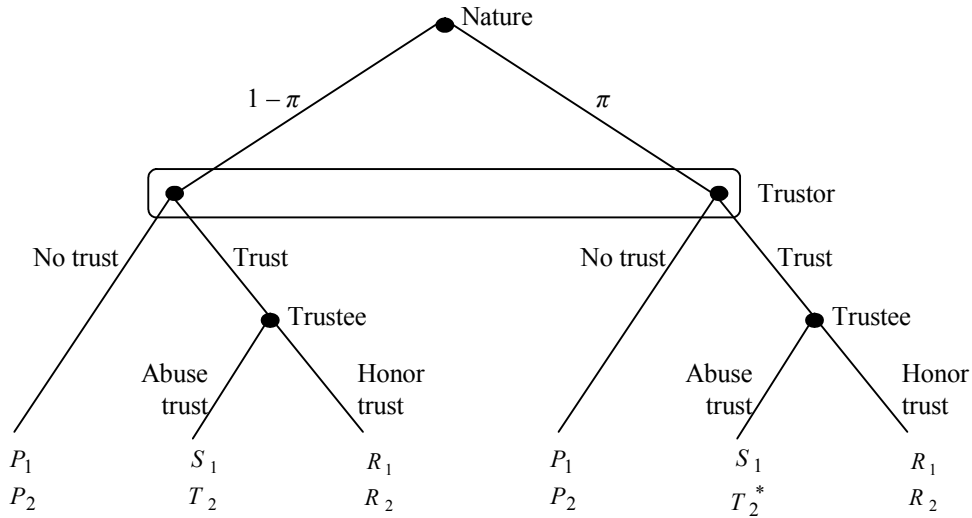


FIGURE 3. Trust Game with Incomplete Information ($S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$, $T_2^* < R_2$)

Social Dilemmas as Games with Pareto-Suboptimal Solutions

Individual rationality in the sense of (subgame-perfect) equilibrium behavior in the Trust Game (for the Trust Game with incomplete information, we focus on the case $\pi < (P_1 - S_1)/(R_1 - S_1)$) and the Investment Game implies that the trustor does not place trust, while the trustee would abuse trust. In the Trust Game, both trustor and trustee are better off if trust is placed and honored compared to the no trust-outcome. In the Investment Game, both trustor and trustee are better off if the trustor trusts the trustee “completely” by sending everything ($M_1 = E_1$) and the trustee honors trust by returning more than has been sent ($K_2 > E_1$), compared to the outcome such that the trustor places no trust at all and sends nothing ($M_1 = 0$).¹⁶ Thus, in both games, individual rationality is at odds with collective rationality in the sense of Pareto optimality. This is the social dilemma property of both games. A natural conceptualization from a game-theoretic perspective is to define a *social dilemma* as a game with a solution that is Pareto-suboptimal. This covers a broad class of games. The Prisoner’s Dilemma and n -person versions of the Prisoner’s Dilemma (e.g., Taylor 1987) are other well-known examples (see, e.g., Harsanyi 1977, 276–80; Dawes 1980; Liebrand 1983; Taylor 1987, chap. 2 for even more examples).¹⁷ Behavior of the actors is often labeled

¹⁵ Analyses of games with incomplete information (see Rasmusen 2007) often involve particularly strong rationality assumptions (e.g., on Bayesian updating) that may be considered questionable as empirical assumptions on human behavior.

¹⁶ See Barrera (2005, 10) for a more detailed discussion of other Pareto-optimal outcomes of the Investment Game and of possible Pareto-improvements of outcomes such that the trustor sends $M_1 < E_1$ and in that sense does not trust the trustee “completely.”

¹⁷ The Trust Game is a one-sided version of the Prisoner’s Dilemma (this is also reflected in our notation for payoffs that is based on standard notation for the Prisoner’s Dilemma). In the Prisoner’s Dilemma, each actor’s equilibrium behavior, defection, can be seen as a form of opportunistic behavior trying to unilaterally exploit the other actor’s cooperation (this is sometimes referred to as defection motivated by “greed” in the literature) as well as a form of protection against the other actor’s defection (sometimes referred to as defection motivated by “fear”). In the Trust Game, only the trustee has an incentive for opportunism, while the trustor’s equilibrium behavior cannot be seen as opportunism but is exclusively protection against the trustee’s opportunism. Of course, modeling economic exchange between a buyer and a seller as a Trust Game with the buyer in the role of

“cooperation” if it overcomes the dilemma in the sense that it is associated with a Pareto-improvement relative to the game-theoretic solution and it provides a Pareto-optimal outcome. It is important to note that the conceptualization of a social dilemma as a game with a Pareto-suboptimal solution is relative to a solution theory for non-cooperative games. A game that is a social dilemma relative to some solution theory need not be a social dilemma under another solution theory. Also, it is not required that the actors’ individually rational strategies in a social dilemma are dominant strategies, i.e., strategies that uniquely maximize the actors’ respective payoffs, no matter what strategies of the other actor(s). Opportunism (“defection”) is a dominant strategy for each actor in the Prisoner’s Dilemma, but the Prisoner’s Dilemma is a special case. The Trust Game and the Investment Game are social dilemmas, while in both games the individually rational strategy of the trustor not to place (any) trust is not a dominant strategy. Even the individually rational strategy of the trustee to abuse trust (completely) is not a dominant strategy in the strict sense as defined above but is only weakly dominant (it is a best reply strategy against all strategies of the trustor but not always a unique best reply). A core feature of conceptualizing a social dilemma as a game with a Pareto-suboptimal solution is the focus on risks of suboptimality due to strategic or opportunistic behavior of at least some actors, rather than suboptimality due to unfavorable external contingencies beyond the control of the actors or suboptimality due to miscoordination (see Van de Rijt and Macy 2006 for a broader characterization of social dilemmas that accounts also for such additional features).

Deriving Testable Hypotheses from Game-Theoretic Models

How to use game-theoretic models for generating empirically testable hypotheses on social dilemmas? For answering this question, one can use Coleman’s (1990, chap. 1) scheme for relating macro- and micro-level propositions in social science explanations (see Figure 4). First, a game’s extensive form models social conditions in the sense of opportunities and restrictions for the actors’ behavior. Thus, the extensive form summarizes the macro-level assumptions related to the top-left node of Coleman’s scheme. Second, the extensive form specifies assumptions on the “independent variables” of rational choice theory, namely, preferences of the actors that are represented via the actors’ payoffs at the endnodes of the game tree and the information actors have when they move during the game. These are the micro-level assumptions related to the bottom-left node of Coleman’s scheme. Also, the extensive form implies the assumptions on macro-micro transitions that are summarized by the vertical arrow 1 in Coleman’s scheme. To see that, note that the extensive form models how an actor’s payoffs and information depend on social conditions. For example, an actor’s payoff function is implied by the extensive form and specifies an actor’s payoff as a function of the strategies of all actors, i.e., as a function of interdependencies.

Thus, the extensive form of a game summarizes empirical assumptions on the macro-level of social conditions, on the micro-level of the actors’ preferences and information, and on macro-micro transitions. Rationality assumptions are micro-level assumptions that are summarized in Coleman’s scheme by arrow 2. These are assumptions on the solution of a game such as that the solution has to be a – possibly “refined” – equilibrium. Game-theoretic analysis then comprises deriving propositions on equilibria of the game and on properties of these equilibria. A simple example – more complex ones will follow – is the analysis of the

the trustor and the seller in the role of the trustee can be misleading because buyer opportunism – such as delayed payment – is abstracted away. Finally, while the actors directly involved in a social dilemma are better off when they manage to overcome the dilemma by realizing Pareto-improvements, this may have detrimental effects for third parties. E.g., collusion between the members of a cartel has bad effects for clients.

Trust Game with incomplete information. An implication of the analysis is that the game has a unique subgame-perfect equilibrium such that trust is placed if $\pi > (P_1 - S_1)/(R_1 - S_1)$. Using game-theoretic rationality assumptions and propositions on equilibria and their properties, one then derives implications on the behavior of rational actors. These implications are represented by the bottom-right node in Coleman's scheme.

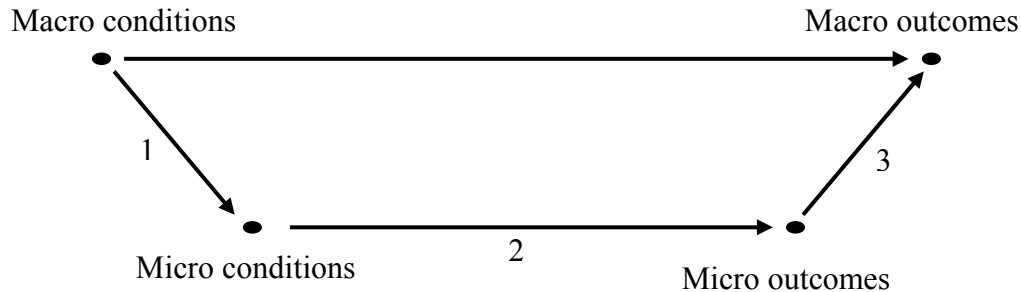


FIGURE 4. Coleman's Scheme

In the next step, one derives implications on how social conditions, changes in social conditions, or differences in social conditions affect the equilibria of the game and, thus, the behavior of rational actors. Deriving such implications often involves comparative statics analyses in one way or the other. For example, the condition $\pi > (P_1 - S_1)/(R_1 - S_1)$ becomes less restrictive if π increases. The probability π of encountering a trustworthy trustee could depend on social conditions that affect processes of socialization and internalization of norms and values. Assuming now that placement of trust becomes more likely when the condition $\pi > (P_1 - S_1)/(R_1 - S_1)$ becomes less restrictive, one can derive testable hypotheses on how social conditions affect the likelihood of placement of trust.¹⁸ In a final step and again using the extensive form of the game, one can derive propositions on macro-level effects. For example, the extensive form of the Trust Game implies that placing and honoring trust is Pareto-optimal. In the specific case of the Trust Game, the macro-level effect "Pareto-(sub)optimality of the outcome of a game" is a rather obvious macro-implication of micro-level individual behavior of trustor and trustee. In other strategic situations, the micro-macro transition of "aggregating" actors' individual behavior into macro-outcomes as indicated by arrow 3 in Coleman's scheme can be much more complex.

TRUST IN ISOLATED ENCOUNTERS

We briefly consider trust in "isolated encounters." Two actors play the Trust Game once and only once. Neither the two actors nor other actors can condition behavior in future interactions on what happens in the Trust Game. From the perspective of the model of atomized interactions on perfect markets of rational and selfish actors, assuming such isolated encounters does already involve a minimal step towards making the atomism assumption more realistic, since interdependencies between trustor and trustee are now taken into account. Isolated encounters are hardly a standard feature of interactions in social and economic life. After all, eBay's feedback forum implies that buyer and seller are *not* involved in an isolated encounter. Hence, isolated encounters are typically studied in the laboratory and are used to study non-standard assumptions on preferences, because other factors such as the

¹⁸ Such an assumption allows for comparative statics analyses and is crucial in deriving testable hypotheses. Strictly speaking, the assumption does not follow from but is used in addition to game-theoretic reasoning. Many empirical applications of game theory, though, do involve similar assumptions explicitly or – more often – implicitly.

social embeddedness can be controlled in the laboratory. More precisely, assume that subjects play the Trust Game from Figure 1, with payoffs $S_1 < P_1 = P_2 < R_1 = R_2 < T_2$ in terms of monetary incentives or points converted into money at the end of the experiment. As explained above, the standard prediction is no trust and, if trust would be placed anyway, it would be abused. For the Investment Game, the analogous prediction is that the trustor sends nothing and, if she would send anything, the trustee would never return anything. Clearly, these are very strong predictions for the behavior of any subject in the laboratory. The predictions are clearly rejected (see Snijders 1996; Snijders and Keren 1999, 2001 on the Trust Game, Berg et al. 1995 on the Investment Game, and Camerer 2003, chap. 2.7 for an extensive review). Similar results are found for other social dilemmas, including two-person Prisoner's Dilemmas (e.g., Sally 1995) and n -person dilemmas (e.g., Ledyard 1995). Camerer and Fehr (2004) provide a useful overview, summarizing results for a series of different interaction situations of the isolated encounters-type, including the social dilemmas mentioned above.

Experiments show that substantial percentages of subjects trust in the Trust Game and send positive amounts in the Investment Game. Also, many subjects in the role of trustee honor trust and return substantial amounts. More generally, opportunism is not ubiquitous in isolated encounters resembling a social dilemma. Different approaches can be envisaged that account for such empirical regularities (see Fehr and Schmidt 2006 for an instructive overview). Each of these approaches involves making the assumptions on the actors more complex in one way or the other. First, one could relax the rationality assumption and employ a *bounded rationality* perspective. For example, one could assume that subjects are used to repeated interactions in life outside the laboratory. As we will see below, placing and honoring trust as well as other forms of cooperative, non-opportunistic behavior can be a result of equilibrium behavior in repeated interactions. The assumption then is that subjects erroneously apply rules in isolated encounters that are appropriate when interactions are repeated (see, e.g., Binmore 1998 for a sophisticated discussion of such approaches). More generally, Binmore (1998, chap. 0.4.2) argues that behavior in experimental games can be expected to be consistent with the assumption of selfish game-theoretic rationality only if the game is easy to understand, adequate incentives are provided, and sufficient time is available for trial-and-error learning (see Kreps 1990b for similar arguments). Second, there are approaches that maintain the rationality assumption but modify the selfishness assumption. These approaches thus abandon the assumption that subjects care exclusively about their own material resources ("utility = own money"). Rather, it is assumed that subjects, or at least some subjects, have *other-regarding preferences*. It is quite often argued (e.g., Fehr and Gintis 2007) that such preferences are due to socialization processes and internalized social norms and values. Also, it is often assumed that subjects may differ with respect to their other-regarding preferences – there may be selfish subjects as well as subjects with other-regarding preferences – and that subjects are incompletely informed on the preferences of other subjects.¹⁹

To get a flavor of how assumptions on other-regarding preferences can be used to account for placing and honoring trust in a Trust Game as an isolated encounter, consider a simple version of a social preferences model, namely, Snijders' (1996; see also Snijders and Keren 1999, 2001) *guilt model*, a simplified version of the Fehr-Schmidt (1999) model of inequity aversion. Assume that actor i 's utility is given by $U_i(x_i, x_j) = x_i - \beta_i \max(x_i - x_j, 0)$ with monetary payoffs x_i and x_j for the actors i and j and $\beta_i \geq 0$ a parameter representing i 's guilt due to an inequitable allocation of monetary payoffs. Hence, in a Trust Game with payoffs in terms of money and $P_1 = P_2$ and $R_1 = R_2$, the trustee's utility from abused trust would be $T_2 -$

¹⁹ Thus, the focus is on a Trust Game with incomplete information similar to the game in Figure 3.

$\beta_2(T_2 - S_1)$, while utilities correspond to own monetary payoffs in all other cases.²⁰ Furthermore, assume actor heterogeneity with respect to the guilt parameter β_i in the sense that there are actors with a large guilt parameter, while β_i is small or even equals zero for other actors, namely, those with selfish preferences. Finally, assume incomplete information of the trustor on the trustee's guilt parameter, with π being the probability that β_2 is "large enough" so that the trustee's utility from abusing trust is smaller than his utility from honoring trust, i.e., $T_2 - \beta_2(T_2 - S_1) < R_2$. "Large enough" thus means $\beta_2 > (T_2 - R_2)/(T_2 - S_1)$, with $(T_2 - R_2)/(T_2 - S_1)$ as a convenient measure of the temptation of an inequity averse trustee to abuse trust. Equilibrium behavior now requires that a trustee with $\beta_2 > (T_2 - R_2)/(T_2 - S_1)$ honors trust, while a trustor places trust if $\pi > (P_1 - S_1)/(R_1 - S_1)$. Employing the logic sketched in subsection on deriving testable hypotheses above, we can assume that placing trust becomes more likely when the condition $\pi > (P_1 - S_1)/(R_1 - S_1)$ becomes less restrictive. Similarly, we can assume that honoring trust becomes more likely when the condition $\beta_2 > (T_2 - R_2)/(T_2 - S_1)$ becomes less restrictive. Furthermore, we can assume that π depends on the trustee's incentives and hence decreases in $(T_2 - R_2)/(T_2 - S_1)$. It follows from this model that the likelihood of placing trust decreases in the trustor's risk $(P_1 - S_1)/(R_1 - S_1)$ as well as in the trustee's temptation $(T_2 - R_2)/(T_2 - S_1)$ and that the likelihood of honoring trust decreases in the trustee's temptation $(T_2 - R_2)/(T_2 - S_1)$. These implications nicely correspond with experimental evidence (see Snijders 1996; Snijders and Keren 1999, 2001).

Obviously, assumptions on other-regarding preferences should be used with care (see, e.g., Camerer 2003, 101; Fehr and Schmidt 2006, 618): (almost) all behavior can be "explained" by assuming the "right" preferences and adjusting the utility function. Thus, one would prefer first of all parsimonious assumptions on other-regarding preferences, adding as few new parameters as possible to the model. Second, when assumptions on other-regarding preferences are employed, one should aim at using the same set of assumptions for explaining behavior in a broad range of different experimental games. Third, one should account not only for well-known empirical regularities but also aim at deriving and testing new predictions. It is therefore important from a methodological perspective that the same set of assumptions on other-regarding preferences is consistent not only with empirical regularities of behavior in Trust Games but also in other social dilemmas, in games involving distribution problems such as the Ultimatum Game (Güth, Schmittberger, and Schwarze 1982; see Camerer 2003 for a survey) or the Dictator Game (Kahneman, Knetsch, and Thaler 1986; see Camerer 2003 for a survey), and in market games. Fehr and Schmidt (1999) argue that their model of inequity aversion succeeds in accounting not only for empirical regularities in a broad class of experimental games but is also consistent with selfish behavior in some settings and non-selfish behavior in others. This is due to heterogeneity between the actors with respect to their inequity aversion. The interaction between actors who are selfish and actors with (stronger) inequity aversion in a setting with incomplete information on other actors' preferences can be a driving force in inducing selfish behavior in settings such as experimental markets and quite some non-selfish behavior in other games, for example, social dilemmas that are isolated encounters (see Shaked 2006; Fehr and Schmidt 2005 for an unusually heated debate on whether the Fehr-Schmidt model is indeed successful in these respects). Finally, it becomes important to empirically discriminate between different assumptions on other-regarding preferences, using careful experimental designs that allow disentangling the different mechanisms assumed in different models of other-regarding preferences (see again Fehr and Schmidt 2006 for a survey).

We refer to Gächter's chapter in this Handbook for a careful discussion of how to refine the model of rational and selfish actors (another useful overview for sociologists is Fehr and

²⁰ Snijders thus neglects that the trustor may derive an additional disutility from abused trust because she envies the inequitable distribution. The Fehr-Schmidt model takes such a disutility into account.

Gintis 2007). In the subsequent sections we return to employing standard assumptions on the level of individual behavior, namely, assumptions on game-theoretic rationality *as well as* basically selfish preferences. We now refine a standard assumption on the social context in neo-classical economics. Rather than assuming atomized interactions – in our case: “isolated encounters” – we explore the implications of “embeddedness” for rational and selfish behavior in social dilemmas.

THEORY AND HYPOTHESES ON EFFECTS OF SOCIAL EMBEDDEDNESS

Roughly, embeddedness (Granovetter 1985) can mean that the actors involved in a focal Trust Game maintain an ongoing relation with prior and expected future interactions. We refer to this as “*dyadic embeddedness*.” An example is the second scenario for the purchase of the antiquarian book where the buyer repeatedly purchases from the same antiquarian. Furthermore, a focal Trust Game can be related to interactions of trustor or trustee with third parties. We refer to this as “*network embeddedness*.” The buyer of the antiquarian book may happen to know others who purchase books from the antiquarian. Finally, there may be institutions that have repercussions for the actors’ opportunities, incentives, or information. We refer to this as “*institutional embeddedness*.” We focus on dyadic and on network embeddedness, also highlighting how institutions can and often do enhance dyadic and network embeddedness. An example is our third scenario for purchasing the antiquarian book where eBay’s feedback forum is an institution that provides network embeddedness for the transaction.

We distinguish two mechanisms, control and learning, through which dyadic and network embeddedness may affect trust. *Control* refers to the case that the trustee has short-term incentives for abusing trust, while some long-term consequences of his behavior in the focal Trust Game depend on behavior of the trustor. More precisely, if the trustee honors trust in the focal Trust Game, the trustor may be able to reward this by applying positive sanctions in the future. Conversely, if the trustee abuses trust in the focal Trust Game, the trustor may be able to punish this by applying negative sanctions. Given dyadic embeddedness, the trustee has to take into account that honoring trust in the focal Trust Game may affect whether or not the trustor places trust again in the future. Given network embeddedness, the trustee has to take into account that a trustor can inform third parties on the trustee’s behavior in the focal Trust Game, such as other trustors with whom the trustee may be involved in future Trust Games. Again, whether or not other trustors are willing to trust the trustee may depend on honoring or abusing trust in the focal Trust Game. Thus, the trustee has to trade off the short-term incentives to abuse trust against the long-term benefits of honoring trust and the long-term costs of abusing trust. This mechanism is also known as conditional cooperation (Taylor 1987) or reciprocity (Gouldner 1960; Blau 1964 [1996]; Diekmann 2004). Reciprocity in this sense (sometimes labeled “weak reciprocity,” e.g., Fehr and Schmidt 2006, 620; Fehr and Gintis 2007) can be driven exclusively by long-term, “enlightened” self-interest of the actors. Thus, reciprocity in this sense differs fundamentally from reciprocal behavior of the trustee in isolated encounters that is based on other-regarding preferences (“strong reciprocity”).

Embeddedness may affect trust through a second mechanism, namely, *learning*. We have already mentioned that the trustor need not be completely informed on the behavioral alternatives and incentives of the trustee. Beliefs of the trustor on the trustee’s characteristics can be affected by information on past interactions. This information can be obtained from past interactions of trustor and trustee, i.e., through dyadic embeddedness. Given network embeddedness, information can also be obtained from third parties who have interacted with the trustee in the past. If a trustee has been trustworthy in past interactions, a trustor might be

more convinced that the trustee will be trustworthy again in the focal Trust Game than if information on untrustworthy behavior of the trustee in the past has been revealed. Table 1 summarizes our distinction between dyadic and network embeddedness as well as between learning and control (Buskens and Raub 2002; see Yamagishi and Yamagishi 1994, 138–39 for a similar discussion of learning and control effects through network embeddedness).

TABLE 1. Types of embeddedness and mechanisms through which embeddedness affects trust

Two mechanisms	Two types of embeddedness	
	Dyad	Network
Control	Sanctioning possibilities of the trustor without involving third parties.	Sanctioning possibilities of the trustor that involve third parties.
Learning	Information about the trustee from past experiences of the trustor.	Information about the trustee from third parties.

Our sketch indicates that embeddedness may help actors to mitigate social dilemmas such as trust problems (see, e.g., Taylor 1987; Kollock 1998 for an overview of different ways in which social dilemmas can be mitigated). Note that embeddedness effects on social and economic interactions and exchange are a common theme of the sociological literature. However, clearly disentangling different types of embeddedness effects and the underlying mechanisms theoretically as well as empirically is often neglected. We now show how game-theoretic tools allow for modeling embeddedness and, more importantly, for deriving hypotheses on effects of embeddedness on trust. This can be done by “embedding” a focal Trust Game in a more complex game. Subsequently, applying the logic for deriving hypotheses from game-theoretic models that has been set out above, conditions are established that promote placing and honoring trust in an embedded Trust Game. Thus, one establishes conditions for an equilibrium of the more complex game, preferably an equilibrium that is a sound candidate for being the game-theoretic solution, such that trust is placed and honored in the focal Trust Game. Propositions on such conditions yield hypotheses on embeddedness effects on trust.

A Simple Model of Dyadic Control: Conditional Trust in the Indefinitely Repeated Trust Game

To see how trust can be a result of purely selfish rational actors who are “enlightened” in the sense that they take long-term effects of their behavior into account, we consider a simple model of control effects through dyadic embeddedness, namely, Kreps’ (1990a) model of a repeated Trust Game (see also Gibbons 2001). In this model, the Trust Game is played repeatedly in rounds $1, 2, \dots, t, \dots$. By way of example, a buyer purchases repeatedly from the same seller of antiquarian books. More precisely, after each round t , another round $t + 1$ is played with probability w ($0 < w < 1$), while the repeated game ends after each round with probability $1 - w$. The focal Trust Game is thus embedded in a more complex game in which the *Trust Game is repeated indefinitely often*. In each round, trustor and trustee observe each other’s behavior. In the repeated game, a strategy is a rule that prescribes an actor’s behavior in each round t as a function of the behavior of both actors in the previous rounds (it is now obvious that a strategy for the repeated game is different from a move in a round of the repeated game). An actor’s expected payoff for the indefinitely repeated Trust Game is the

discounted sum of the actor's payoffs in each round, with the continuation probability w as discount parameter. For example, a trustor who places trust throughout the repeated game, with trust being honored throughout, receives payoff $R_1 + wR_1 + \dots + w^{t-1}R_1 + \dots = R_1/(1 - w)$. Thus, using Axelrod's (1984) apt label, the continuation probability w represents the "shadow of the future": the larger w , the more an actor's payoff from the repeated game depends on what the actor receives in future rounds.

In the indefinitely repeated Trust Game, the trustor can exercise control. She can use a conditional strategy that rewards a trustee who honors trust in a focal Trust Game by placing trust again in future games. Conversely, a conditional strategy of the trustor can punish abuse of trust by the trustee in the focal Trust Game through not placing trust in at least some future games. Other forms of rewards and punishment are excluded in this simple scenario.²¹

If the trustor uses weak reciprocity in the sense of implementing a conditional strategy, the trustee can gain T_2 rather than R_2 in the current Trust Game by abusing trust. However, abusing trust will then be associated with obtaining only P_2 in (some) future encounters with no trust placed by the trustor, while honoring trust will result in larger payoffs than P_2 in those future encounters if the trustor goes on placing trust. Moreover, the larger the shadow of the future, the more important are the long-term effects of present behavior. Thus, anticipating that the trustor may use a conditional strategy, the trustee has to balance short-term ($T_2 - R_2$) and long-term ($R_2 - P_2$) incentives. It can be shown that weak reciprocity can be a basis for rational trust in the sense that the indefinitely repeated Trust Game has an equilibrium such that trust is placed and honored in each round. Consider the strategy of the trustor that is associated with the largest rewards for trustworthy behavior of the trustee and with the most severe sanctions for untrustworthy behavior. This is the strategy that prescribes to place trust in the first round and also in future rounds, as long as trust has been placed and honored in all previous rounds. However, as soon as trust is not placed or abused in some round, the trustor refuses to place trust in any future round. Such a strategy is often labeled a "*trigger strategy*" because deviation of the trustee from the "prescribed" pattern of behavior triggers a change in the trustor's behavior. Straightforward analysis shows that always honoring trust (and always abusing trust as soon as there has been any deviation from the pattern "place and honor trust") is a best reply of the trustee against a trigger strategy of the trustor if and only if

$$(1) w \geq (T_2 - R_2)/(T_2 - P_2).$$

This condition requires that the shadow of the future is large enough compared to $(T_2 - R_2)/(T_2 - P_2)$, a convenient measure for a selfish trustee's temptation to abuse trust. The condition refers exclusively to the incentives of the trustee and not at all to the incentives of the trustor.²² This highlights that placing and honoring trust in the indefinitely repeated Trust Game is driven by the strategic interdependence of the actors.

²¹ Obviously, one could model reward and punishment options in other ways. In the Trust Game, one could add options for "direct" sanctions by the trustor after the trustee has honored or, respectively, abused trust rather than sanctions through behavior of the trustor in future games. See, e.g., Gächter's chapter in this Handbook and Fehr and Gintis (2007) for overviews of models that include such sanction possibilities in a focal social dilemma game itself and empirical evidence indicating that such sanction possibilities affect behavior in non-repeated social dilemmas quite dramatically. Again, models assuming other-regarding preferences can be used to account for those effects in non-repeated games. In the indefinitely repeated Trust Game, in contrast, assumptions on other-regarding preferences are not needed.

²² In contrast, the expression derived for the temptation in isolated encounters using guilt incorporates also the trustor's payoff S_1 .

If condition (1) applies, the indefinitely repeated Trust Game has an equilibrium such that trust is always placed and honored. This equilibrium is likewise subgame-perfect.²³ This implies that the trustor's (implicit) promise to reward trustworthy behavior of the trustee by placing trust again in the future and her (implicit) threat to punish abuse of trust by not placing trust again are credible. Enlightened self-interest can thus be a basis for trust among rational actors in the sense of placing and honoring trust being equilibrium behavior.²⁴ The equilibrium, however, is not unique. For example, never placing trust, while placed trust would always be abused is always an equilibrium of the indefinitely repeated game. The "folk theorem" (e.g., Fudenberg and Maskin 1986; Rasmusen 2007, chap. 5.2) for repeated games implies that the indefinitely repeated Trust Game has many other equilibria, too, for large enough w . Thus, an equilibrium selection problem emerges. A typical, though sometimes implicit, argument in the literature on equilibrium selection in this context is "payoff dominance." An equilibrium is payoff dominated if there is another equilibrium that is associated with higher payoffs for at least some actor and is not associated with lower payoffs for any actor. In the indefinitely repeated Trust Game, an equilibrium that implies placed and honored trust throughout the game is evidently not payoff dominated by other equilibria, while the no-trust-throughout equilibrium is payoff dominated. Also, one could argue that the equilibrium selection problem for the indefinitely repeated Trust Game highlights that communication could help rational actors to coordinate on the trigger strategy equilibrium while, theoretically, communication would be expected to be irrelevant in a Trust Game as an isolated encounter between rational and selfish actors.²⁵

It can be argued that the trigger strategy equilibrium for the indefinitely repeated Trust Game is implausible. For example, the equilibrium implies that trust is *always* placed and honored, while one might rather expect less than "perfect" trust, even under favorable conditions for trust. One can show (see, e.g., Taylor 1987) that there are also other equilibria that induce placement of trust and honoring trust only in some, rather than in all rounds of the game, while this pattern of behavior is again backed by a variant of the trigger strategy: as soon as there is a deviation from the "prescribed" pattern, the trustor never places trust again. Then, however, the problem becomes even more severe to select one out of a wealth of such equilibria as a "solution candidate." Moreover, although no deviations occur in equilibrium, it may seem implausible that trust *would* break down *completely* after the first deviation. A complete breakdown of all future trust may seem implausible for a single deviation from the trigger strategy equilibrium in the sense of honored trust in each round of the game as well as for a single deviation from other patterns of behavior that are backed by a variant of a trigger strategy. This counter-intuitive feature can be circumvented, for example, by considering a game with imperfect monitoring (e.g., Green and Porter 1984). Assume that the trustor, after placing trust, cannot observe the trustee's behavior, but can only observe the outcome of that behavior. This outcome, in turns, depends on the trustee's behavior but also on chance: a low payoff for the trustor after placement of trust can be due to abuse of trust by the trustee but

²³ This shows that rational actors may cooperate in a non-cooperative game (see note 4). It should be observed that the trigger strategy is not the only conditional strategy that can be used to stabilize trust and trustworthiness as a result of equilibrium behavior of rational actors. Other conditional strategies that use less severe punishments than the trigger strategy can do so, too. However, one then needs further equilibrium conditions rather than exclusively condition (1).

²⁴ Coleman, in his meanwhile classic sketch, clearly intuited this result when he argued that an important feature of socialization is "coming to see the long-term consequences to oneself of particular strategies of action" rather than the internalization of norms (1964, 180). Voss (1982) seems to be the first sociologist who realized explicitly that the theory of repeated games has important implications for the problem of order and cooperation in social dilemmas.

²⁵ See, e.g., Bohnet (1997) for experimental evidence that, empirically, communication does affect behavior also in social dilemmas as isolated encounters.

can also be due to “bad luck.” Such a scenario is much more difficult to analyze. The trustor now has to solve an “optimal punishment” problem. If the trustor never punishes, or applies too lenient punishments, a rational trustee would always abuse trust. But too severe punishments imply more than necessary (in terms of deterring the trustee from abusing trust) losses for trustor and trustee. Equilibrium behavior that generates some honored trust throughout the game now requires that the trustor punishes the trustee by placing no trust occasionally rather than eternally.

Even though the trigger strategy equilibrium for the indefinitely repeated Trust Game may have implausible features, consider an interpretation of the equilibrium condition that follows the logic for deriving testable hypotheses from game-theoretic models as set out in section on deriving testable hypothesis above. Rather than claiming that actors indeed use trigger strategies, one proceeds from the observation that condition (1) is a necessary and sufficient condition for equilibria in the indefinitely repeated Trust Game such that trust is placed and honored throughout the game. One could then again assume that placing and honoring trust becomes more likely when the condition becomes less restrictive. This leads directly to testable hypotheses on control effects through dyadic embeddedness. Specifically, one would expect that the likelihood of placing and honoring trust increases in the shadow of the future w and decreases in the temptation $(T_2 - R_2)/(T_2 - P_2)$ for a selfish trustee.

The results for the indefinitely repeated Trust Game can be generalized. For example, analogous results hold for an indefinitely repeated Investment Game. Friedman (1971, 1990) shows that analogous results apply to a broad class of indefinitely repeated 2- and n -person games. Roughly speaking, if a social dilemma is repeated indefinitely often and the shadow of the future is large enough relative to the short-term incentives of the actors, there exists an equilibrium of the indefinitely repeated game such that the actors cooperate: the equilibrium of the repeated game induces a Pareto-optimal outcome and a Pareto-improvement compared to the Pareto-suboptimal solution of the original dilemma. Of course, these generalizations should be interpreted with care. For example, trigger strategies require the observability of the behavior of other actors. Hence, the underlying assumption that each actor receives reliable information on each other actor’s behavior in each round of the game is crucial, while such an assumption will often be rather problematic from an empirical perspective in games with many actors (see, e.g., Bendor and Mookherjee 1987).

Network Control

Models of repeated Trust Games can be extended to account for control effects due to network embeddedness in addition to dyadic embeddedness. In these extended models, the trustee interacts with a set of trustors, while the trustors are connected through a network that allows for communication about the behavior of the trustee. The focal Trust Game is now embedded in a more complex game that comprises Trust Games of the trustee with different trustors. The important feature is that the trustor in a focal Trust Game can transmit information on the trustee’s behavior in that game to other trustors. Next to direct reciprocity exercised by the trustor who interacts herself with the trustee in the focal Trust Game, network embeddedness allows for indirect reciprocity exercised by other trustors. A trustee contemplating to honor or abuse trust in a focal Trust Game now has to consider future sanctions by the trustor with whom he interacts in the focal Trust Game as well as sanctions that can be applied by other future trustors who receive information on the trustee’s behavior in the focal Trust Game and who may condition their future behavior on that information. In terms of our example of purchasing antiquarian books, we thus now consider variants of an eBay feedback forum.

First, such network embeddedness can be a substitute for dyadic embeddedness (see Kreps 1990a, 106–8). Assume that the trustee interacts with a different trustor in each round of the indefinitely repeated Trust Game in the previous section. Thus, each trustor plays the Trust Game only once with the trustee. Dyadic embeddedness is then removed completely from the repeated game and has been replaced by network embeddedness. However, if the trustor in a given round is reliably informed on what has happened in previous rounds, each trustor can condition his behavior in a given round in the same way as a trustor who plays in each round and uses a trigger strategy: trust is placed if and only if there is no information that trust has not been honored before. Evidently, the trustee’s best reply against such behavior of the trustors is again to honor trust in each round if condition (1) is fulfilled. Conversely, indeed placing trust is then best reply behavior also of the trustors. Hence, we see that network embeddedness can induce trust among rational and selfish actors.

Dyadic as well as network embeddedness is included in more complex models (e.g., Weesie, Buskens, and Raub 1998; Buskens and Weesie 2000a; Buskens 2002, chap. 3; see Raub and Weesie 1990 for a related model of network embeddedness for the Prisoner’s Dilemma). In these models, a trustee interacts with a trustor in an indefinitely repeated Trust Game. After the interaction with a given trustor ends, the trustee goes on playing an indefinitely repeated Trust Game with another trustor, while information on behavior in the Trust Games with the first trustor is communicated to the second trustor with some probability. Interactions with a third trustor start after the interactions with the second trustor ended and so forth. These models are relatively general and allow for quite some heterogeneity with respect to various features: the incentive T_2 for abusing trust varies between games, the probability of starting interactions with the trustee as well as the continuation probability for these interactions varies between trustors, and the probability of information transmission varies between pairs of trustors. One can then study subgame-perfect equilibria such that trustors place trust if T_2 is not “too large” and if they do not have information that trust has ever been abused. Trust is thus backed up by variants of trigger strategies. A nice feature of these models is that they account for the intuition that trust will not always be placed. In addition to hypotheses on how the likelihood of trust is affected by the shadow of the future and the short-term incentives of the trustee, such models allow for deriving hypotheses on effects of network characteristics. Specifically, the likelihood of placing and honoring trust in a focal Trust Game increases in the density of the network of trustors as well as in the trustor’s outdegree, i.e., her probability to transmit information to the next trustor who interacts with the trustee. This is intuitively plausible since network density as well as outdegree increase the sanction possibilities of the trustor. Hence, if the trustee considers the long-term consequences of his behavior, higher network density and outdegree allow for placing and honoring trust even if the trustee’s short-term incentive to abuse trust is fairly large.

A major problem of these models is that they assume that information is reliable and that incentive problems associated with the supply of information are neglected (see, e.g., Lorenz 1988; Raub and Weesie 1990, 648; Williamson 1996, 153–55; Blumberg 1997, 208–10; Buskens 2002, 18–20). However, supplying information on the trustee’s behavior is a contribution to a public good, namely, enforcing trustworthy behavior of the trustee. Such contributions are problematic: after all, public good production is itself a social dilemma when contributions are costly (this feature is often discussed as a major problem of institutions such as eBay’s feedback forum; see, e.g., Bolton and Ockenfels 2006). Moreover, information from third parties can be inconsistent with own experiences. Also, information from third parties can be problematic due to misunderstanding or strategic misrepresentation: imagine that the trustors are competitors who purchase the same goods from the same seller. In a nutshell, one would expect that effects of network embeddedness are attenuated when

such problems become more serious. Notice, too, that we have focused on the case of network control in the sense that other trustors can sanction the trustee in future interactions. This is control through “voice” in Hirschman’s (1970) sense. A different case of network control is that a trustor has access to alternative trustees and can exercise control through “exit”: whether or not the trustor interacts again with the trustee in the future depends on the trustee’s behavior in the focal Trust Game. Modeling network control through exit opportunities for the trustor is not trivial (see Hirshleifer and Rasmusen 1989; Schüßler 1989; Vanberg and Congleton 1992 for related models) but one would expect in general that the likelihood of placing and honoring trust increases in the trustor’s exit opportunities.

Game-Theoretic Models of Trust Based on Learning and Control

The game-theoretic models of embeddedness effects on trust discussed above have been (repeated) games with complete information: roughly speaking, each actor is informed on the behavioral alternatives and incentives of all actors. Specifically, trustors are completely informed on the behavioral alternatives and the incentives of the trustee. Hence, there is no need – and no opportunity – for trustors to learn during the game about unobservable characteristics of the trustee. This means that these models do not yield hypotheses on learning effects of embeddedness.²⁶

Hypotheses on control as well as learning effects can be derived from models of games with incomplete information. Typically, these are models of finitely repeated games. To get a flavor of these games, consider first of all a finitely repeated game with complete information. Assume that trustor and trustee play the Trust Game from Figure 1 repeatedly, namely, N times. Clearly, in the final round, equilibrium behavior requires that trust would be abused and no trust will be placed. However, this means that behavior in the last but one round cannot have effects on behavior in the final round. Hence, no trust will be placed in the last but one round and so forth, back to the first round. This is the famous backward induction argument showing that placing and honoring trust cannot be a result of rational and selfish behavior in a finitely repeated Trust Game with complete information.

Things change dramatically by introducing *incomplete information in the finitely repeated Trust Game* (see Camerer and Weigelt 1988; Dasgupta 1988; Neral and Ochs 1992; Bower, Garber, and Watson 1997; Buskens 2003). Introducing incomplete information means relaxing another core assumption of the standard rational choice model. As in the subsection on the Trust Game with incomplete information, assume that there is a positive ex-ante probability π that the trustee actually has no incentive to abuse trust, i.e., his payoff from abusing trust is $T_2^* < R_2$ (an alternative assumption leading to essentially the same results would be to assume that the trustee has no opportunity to abuse trust with probability π). The trustor knows the probability π but cannot directly observe whether the trustee’s payoff from abusing trust is T_2 or T_2^* . Now, if the trustor places trust in some round of the repeated game that is not the final round, trust may be honored for one of two very different reasons. First, the trustee’s payoff could be $T_2^* < R_2$ so that there is no incentive at all for the trustee to abuse trust. Second, the trustee’s payoff could be $T_2 > R_2$ but the trustee follows an incentive for reputation building. The trustee knows that if he abuses trust, the trustor can infer for sure that the trustee’s payoff from abusing trust is $T_2 > R_2$ and will thus never place trust again in future rounds. On the other hand, if the trustee honors trust, the trustor remains uncertain about the

²⁶ One might argue that learning is still possible in these models, since there are many equilibria and it is not clear why actors should choose the same equilibrium to start with. We disregard this issue, assuming that actors coordinate instantly on the same equilibrium (see, e.g., Fudenberg and Levine 1998, 20).

trustee's incentives and may place trust again in the future. Conversely, the trustor can anticipate on such behavior of the trustee and may therefore be inclined to indeed place trust. In this game, the trustor can control the trustee in that placing trust in future rounds depends on honoring trust in the current round and the trustor can learn about the incentives of the trustee from the trustee's behavior in previous rounds. The result is a subtle interplay of a trustor who tries to learn about and to control the trustee, taking the trustee's incentives for reputation building into account, and a trustee who balances the long-term effects of his reputation and the short-term incentives for abusing trust, taking into account that the trustor anticipates on this balancing.

It can be shown that the game has a sequential equilibrium (a further refinement of the subgame-perfect equilibrium concept that can be applied to games with incomplete information, Kreps and Wilson 1982) that does involve placing and honoring trust in some rounds of the repeated game. More precisely, in that equilibrium, the game starts with trust being placed and honored in a number of rounds. Afterwards, a second phase follows in which the trustor and the trustee with $T_2 > R_2$ randomize their behavior until the trustor does not place trust or the trustee abuses trust. After trust has not been placed or has been abused for the first time, the third and last phase starts in which no trust is placed until the end of the game. A remarkable feature of the model is that quite some honored trust can be induced by equilibrium behavior even if the probability π that the trustee has no incentive to abuse trust is small. In the equilibrium, learning occurs – in the sense that the trustor updates her belief about the probability that she is playing with a trustee without an incentive to abuse trust – if trust is abused and in the second phase as long as trust is honored. Learning is rational in the sense of Bayesian updating. The first phase of the game with trust being placed and honored is shorter, the higher the risk $(P_1 - S_1)/(R_1 - S_1)$ for the trustor, the smaller the number of rounds of the repeated game, and the smaller the ex-ante probability π . While the risk for the trustor is a driving force of the model, the trustee's temptation $(T_2 - R_2)/(T_2 - P_2)$ only affects behavior in the second randomization phase of the repeated game. Quite counterintuitively, the probability that the trustor places trust in that phase increases (!) in the trustee's temptation (see Buskens 2003, 239 for an explanation).

Game-theoretic models with incomplete information such as the finitely repeated Trust Game are complex and not easily analyzed. They become even more complex by including learning due to network embeddedness. A shortcut linking learning effects of network embeddedness to such models would be to assume that the trustor's ex-ante probability π of interacting with a trustee who would never abuse depends on information the trustor receives from third parties such as other trustors who played Trust Games previously with the trustee. Specifically, based on information diffusion models in networks of trustors (e.g., Buskens 2002, chap. 4) and assuming that the information about the trustee is positive (it is information that the trustee has honored rather than abused trust), one would expect that the ex-ante probability π increases in the density of the network of trustors as well as in the extent to which the trustor in the focal Trust Game receives information about the trustee from other trustors, i.e., increases in the trustor's indegree.

A more explicit game-theoretic model of network effects in games with incomplete information has been provided by Buskens (2003). In that model, the trustee plays Trust Games with two different trustors A and B . With some probability, each trustor can inform the other trustor on the trustee's previous behavior. We can conceive of the probability that trustor A transmits information to trustor B as A 's outdegree and B 's indegree (and vice versa). Thus, trustor A controls the trustee through her outdegree and learns from B about the trustee through her indegree. If each trustor transmits information to and receives information on the trustee from the other trustor with sufficiently high probability, the first phase of the repeated game such that trust is placed and honored becomes longer and in this sense network

embeddedness increases trust. Counterintuitively, in the randomization phase, the probability of placing trust decreases (!) in the probabilities of information transmission.

Summarizing and interpreting the results of the game-theoretic models for learning effects through dyadic and network embeddedness in line with the approach outlined in the game-theory section yields the hypotheses that the likelihood of placing and honoring trust decreases in the trustor's risk $(P_1 - S_1)/(R_1 - S_1)$ and increases if the trustor's previous experiences with the trustee are positive (the trustee honored trust) rather than negative (the trustee abused trust). Furthermore, assuming that the trustor receives positive information about the trustee from other trustors, the likelihood of placing and honoring trust increases in the density of the network of trustors, and in the trustor's indegree. These are relatively robust hypotheses that lend themselves also for empirical research outside the laboratory. The counterintuitive hypotheses on behavior in the randomization phase of the games are clearly best tested in lab experiments.

Models for control and learning effects of embeddedness in games with incomplete information are not only problematic in that they use very strong assumptions on the actors' rationality (in the sense of sequential equilibrium), including rational (Bayesian) updating of beliefs. These models are also problematic in that they neglect learning on other features than unobservable characteristics of the trustee. For example, a trustor could try to use information she receives from other trustors for inferring how to reasonably cope with trust problems. Also, past interactions may give rise to other effects than exclusively learning. For example, actors may have pledged investments in their relation through past interactions and these investments affect the incentives in the focal Trust Game.

The attractive feature of game-theoretic models involving incomplete information is that control and learning can be analyzed simultaneously. The price tag attached to these models is a set of rather strong assumptions on the actors' rationality. Alternatives are "pure" learning models in which actors adapt their behavior based on past experiences. Actors try to optimize short-term outcomes, while not (or "hardly") looking ahead. This implies, too, that actors do not take other actors' incentives into account (see, e.g., Camerer 2003, chap. 6 for a useful overview of learning models; Macy and Flache 1995, 2002; Flache and Macy 2002 provide applications to social dilemmas; Buskens 2002, chap. 4 is an example of a model of learning in networks). Hence, these models neglect control effects. Typically (see Buskens and Raub 2002, 173–76) learning models yield hypotheses that the likelihood of placing trust decreases in the trustor's risk $(P_1 - S_1)/(R_1 - S_1)$. Also, the trustor's estimation of the probability π that trust will be honored will typically increase with positive information about the trustee's behavior in previous interactions, be it information from the trustor's own previous interactions with the trustee or information from third parties. Therefore, one would again hypothesize that more positive information increases the likelihood of placing trust.

Institutional Embeddedness

We have already mentioned that institutions often enhance dyadic and, specifically, network embeddedness by allowing actors to inform other actors, thus enhancing opportunities to exercise control, as well as to receive information from other actors, thus enhancing learning opportunities (see Greif's chapter in this Handbook for a survey of rational choice research on institutions). Modern examples are eBay's feedback forum and similar reputation systems used in the Internet economy. An institution such as the eBay feedback forum allows buyers to evaluate sellers and to collect information on sellers from other buyers. Similarly, sellers can provide and receive feedback on buyers. Fascinating cases of similar institutions in medieval trade are the Maghribi traders' coalition (Greif 1989; see

2006 for a comprehensive treatment) and the law merchants (Milgrom, North, and Weingast 1990; see also Klein 1997 for more examples and Schramm and Taube 2003 for the more recent example of the Islamic *hawala* financial system). While they help actors to overcome social dilemmas in economic exchange, such institutions cannot be taken for granted, for example, due to incentive problems associated with the provision of (correct) information and feedback. Hence, a strong feature of the models provided by Greif as well as Milgrom et al. is that the institutions are “endogenized” in the sense that it is shown that they are themselves result of equilibrium behavior in repeated games.

We briefly mention another type of institutions that help actors in overcoming trust problems and other social dilemmas. Contract law and other institutions often provide opportunities for actors to modify themselves their own (future) incentives or, as Coleman (1990) put it, to construct their social environment. Actors do so by incurring *commitments* (Schelling 1960; Williamson 1985). For example, a seller in the role of the trustee voluntarily provides a guarantee before the Trust Game itself is played. The guarantee modifies the subsequent incentives for trustor and trustee in the Trust Game. Commitments such as guarantees can promote trust by reducing the trustee’s incentive for abusing trust, by providing compensation for the trustor in case trust is abused, or by signaling that the trustee will not (or cannot) abuse trust. Game-theoretic models can be used to specify conditions such that commitments are incurred and induce placing and honoring trust (e.g., Weesie and Raub 1996; Raub 2004). These models allow for deriving hypotheses on how characteristics of the commitment such as the (transaction) costs associated with incurring a commitment, the size of the reduction of the trustee’s incentive to abuse trust, or the size of the compensation for the trustor in case of abused trust affect the likelihood of incurring a commitment as well as the likelihood of placing and honoring trust. In these models, a context that provides opportunities for incurring a commitment is assumed as exogenous. The commitment itself can then be conceived as a “private institution,” voluntarily created by the actors involved in a social dilemma for overcoming the dilemma. A strength of the models is then again that the private institution is not taken for granted but is itself an outcome of equilibrium behavior.

Institutional embeddedness can be a substitute as well as a complement for dyadic and network embeddedness. Given institutional embeddedness, actors can overcome trust problems and other social dilemmas even if dyadic and network embeddedness are absent or are insufficient to promote trust, for example, due to large incentives for abusing trust (“golden opportunities”). Also, some models are meanwhile available that study effects of dyadic, network, and institutional embeddedness on trust simultaneously (e.g., Weesie et al. 1998). Table 2 summarizes the hypotheses discussed in this section.

TABLE 2. Hypotheses on effects of social embeddedness

Two mechanisms	Two types of embeddedness	
	Dyad	Network
Control	1. Trust and trustworthiness decrease with the temptation to abuse trust for the trustee and increase with the likelihood that an interaction is repeated.	3. Trust and trustworthiness increase with the density of the trustor's network, her outdegree, and the availability of institutions that provide information.
Learning	2. Trust and trustworthiness decrease with the trustor's risk and increase with positive experiences with a trustee.	4. Trust and trustworthiness increase with the density of the trustor's network, her indegree, and the availability of institutions that provide information (given that information about the trustee is predominantly positive).

EMPIRICAL RESEARCH ON EFFECTS OF SOCIAL EMBEDDEDNESS

We organize our overview of empirical evidence on effects of social embeddedness by type of research design, focusing on evidence closely related to the hypotheses summarized in Table 2. First, lab experiments are used for testing hypotheses on embedded effects (see Cook and Cooper 2003 for an overview of experimental studies on how other elements in the social context can affect trust). Experiments allow for control over the variation in important independent variables and the causal relation between manipulations and outcome differences is mostly obvious. The disadvantage is that set-ups are often rather artificial. Subjects are typically students who are engaged in abstract interactions. This questions external validity. Therefore, evidence from settings beyond the lab is a complement to experimental evidence. In the second part, we thus review evidence from survey studies, with some emphasis on evidence from on-line transactions. It is typically difficult to disentangle learning and control effects of embeddedness in these studies. We conclude our overview with a brief sketch of two vignette experiments that were specifically designed to overcome this problem. Clearly, vignette experiments have their limitations, too. E.g., incentives for subjects are problematic, since vignette designs involve hypothetical decisions in hypothetical situations. One may thus conclude that it makes sense to employ different and complementary research designs, each having specific strengths and shortcomings, for testing hypotheses on embeddedness effects in order to assess the robustness of empirical findings.

Laboratory Experiments

Effects of dyadic embeddedness

Camerer and Weigelt (1988) initiated experimental research that aims at carefully testing hypotheses on behavior in *finitely repeated Trust Games* with incomplete information, with follow-up studies by Neral and Ochs (1992), Anderhub, Engelmann, and Güth (2002), and Brandts and Figueras (2003). See Camerer (2003, 446–53) for a more detailed overview of

these experiments.²⁷ While experiments on one-shot Trust Games focus on payoff effects and reveal that these effects, in particular effects of risk and temptation, are strong, experiments on repeated Trust Games focus on embeddedness effects. Experiments confirm that trust as well as trustworthiness are high in early rounds and decrease when the end of the repeated game approaches (dyadic control). Trust is almost absent after any abuse of trust, while trust remains relatively high as long as trust has been honored (dyadic learning). However, the trustor's tendency to place trust as long as trust has been honored does not increase as the end of the game comes nearer. This is consistent with the theory, since trustors have to realize that trustees with an incentive to abuse trust also have an incentive to make trustors believe that they do not abuse trust, while these trustees will in fact abuse trust toward the end of the game. Brandts and Figueras (2003) also find that trust and trustworthiness increase with the probability that a trustee has no incentive to abuse trust. Summarizing, the sequential equilibrium predicts quite some global patterns of behavior reasonably well. However, the experiments of Neral and Ochs (1992), Anderhub et al. (2002), and Brandts and Figueras (2003), also show that behavior of subjects does not completely follow the predicted sequential equilibria. For example, it is predicted that in the second phase of the game in which trustors and trustees with an incentive to abuse trust both randomize, the probability that trustors trust increases (!) with the temptation for the trustee to abuse trust. This implication is not only counterintuitive but also inconsistent with experimental findings.

Results from some *other experiments* are quite in line with these findings. Gauthschi (2000) reports findings for finitely repeated Trust Games that comprise two or three rounds of play. He finds that positive past experience matters (dyadic learning) and that the number of remaining rounds to be played also increases trust (dyadic control). For a more contextualized setting with buyers and sellers and an incentive structure similar to the Trust Game, Kollock (1994) finds similar evidence. Still, the studies by Gauthschi and Kollock report quite some untrustworthy behavior by trustees very early in the games. This can be explained by the difference that subjects in the studies of Gauthschi (2000) as well as Kollock (1994) play relatively few games, while subjects in the studies by Camerer and Weigelt (1988) and the related follow-up studies play very many games. Camerer and Weigelt (1988) as well as their followers mainly analyze the later games in the experiment. In this way, suboptimal behavior is minimized. For example, if trustees build up more experience they "learn" that they actually earn less if they behave opportunistically too early. In the studies by Gauthschi (2000) and Kollock (1994), subjects have much less opportunities for this type of learning.

Engle-Warnick and Slonim (2004, 2006) compare *finitely* and *indefinitely repeated games*. In principle, the trustor's opportunities to exercise control in an indefinitely repeated game with constant continuation probability are the same in round t and in round $t+1$. Still, the authors find decreasing trust over time in such games. However, this decrease is much smoother than in the finitely repeated games. This can be understood in the sense that learning effects in terms of negative experiences reduce trust over time and subsequently trust seems to be difficult to restore. On the other hand, trust remains reasonably high because control opportunities do not diminish over time and enable some pairs to continue to trust each other. An additional explanation for decreasing trust in indefinitely repeated games might be that subjects believe that after many repetitions of the game the probability increases that a specific round will be the last one, even if experimenters do their very best to make it apparent that the continuation after every round is constant (e.g., by using a publicly thrown dice).

²⁷ There are sizeable parallel literatures on experiments with repeated social dilemmas like the Prisoner's Dilemma (overviews: Dawes 1980; Colman 1982, chap. 7; Sally 1995 for overviews), public good games (overview: Ledyard 1995), and still other strategic interactions (overview: Camerer 2003).

While there are many experiments on the *Investment Game*, only few use the finitely repeated Investment Game. The findings of Cochard, Nguyen Van, and Willinger (2004) are in line with empirical regularities that have been found for the Trust Game. Subjects send more in the Investment Game if there is a longer future ahead (dyadic control), but if receivers do not return enough they stop sending (dyadic learning). In early rounds, trustors send more if the trustees return more. While Cochard et al. refer to this finding as a reciprocity effect, it can also be interpreted as a learning effect. Again, there is a strong endgame effect although it is observed very late in the games. Trustees start to return less in the last-but-one round. Trustors react on low return rates by sending less in the last round, but there is no significant evidence that trustors send less as a pure result of being in the last round.²⁸

Effects of network embeddedness

Experiments with Trust Games or Investment Games that include network embeddedness are still rare. Bolton, Katok, and Ockenfels (2004) compare one-shot Trust Games that are isolated encounters in the strict sense, finitely repeated Trust Games (with the same partner), and a third treatment in which subjects play multiple one-shot Trust Games with different partners, but obtain information about the past behavior of their partners in interactions with other subjects (for a similar set-up and results, see Bohnet and Huck 2004). In the one-shot Trust Games, trust and trustworthiness reduce quickly after subjects have some experience. Trust and trustworthiness remain high in the repeated Trust Games and collapse only in the last couple of rounds. This finding resonates with evidence on effects of dyadic embeddedness. In the third treatment, there is initially less trust and trustworthiness than in the finitely repeated Trust Game setting, but trustees apparently learn fast enough that they have a considerable problem if they do not honor trust. In this treatment, trust and trustworthiness stabilize for some time in the middle of the series of interactions, although at a somewhat lower level than in the repeated Trust Game setting. Again, trust collapses in the last few rounds. Bolton et al. (2004) interpret their third treatment as an experimental implementation of an institutionalized reputation system that is common for on-line transactions. The treatment could also be interpreted as a complete network in which information diffusion is perfect. Below we will come back to this. Note that the Bolton et al. reputation treatment involves opportunities for learning as well as control through third parties. While indicating that network embeddedness matters, it remains unclear to which mechanism – learning or control or both – trust can be attributed.

Barrera and Buskens (2006) introduce a network setting with subjects playing finitely repeated Investment Games in groups of six. There are two trustees and four trustors. The four trustors play in pairs with the same trustee and we vary the extent of information trustors obtain about transactions from the trustors who plays with the same trustee and from trustors that play with the other trustee. It turns out that there is not much variation in the level of amounts sent depending on the amount of available information in the network. Similar to findings from other experiments, within dyads, trustors send more after positive experiences with the trustee and trust as well as trustworthiness collapse near the end of the game. This is once again evidence for dyadic control as well as dyadic learning. In addition, trustors send more if they observe that the other trustor who plays with the same trustee also sent more, which provides evidence for network learning. Surprisingly, this effect is only weakly related to the amount the trustee returns to the other trustor.²⁹ Buskens (2003) implies that with

²⁸ Experiments employing the closely related gift-exchange game (Van der Heijden et al. 2001; Gächter and Falk 2002) likewise show that repeated play increases the efficiency of outcomes and that endgame effects occur in a similar manner.

²⁹ The effects of how much another trustor sends and how much is returned to this other trustor are difficult to disentangle for two reasons. First, trustees are often rather consistent in how much they return to one or the other

increasing network embeddedness the sequential equilibrium collapses later. Barrera and Buskens (2006) check for this effect, controlling for learning experiences, but they do not find evidence for this network control effect.

Survey Studies

There is quite some qualitative evidence that dyadic embeddedness (e.g., Uzzi 1996) and network embeddedness (e.g., Wechsberg 1966) affect trust. However, we focus on more quantitative evidence from surveys. As we will see, although many surveys offer evidence for effects of embeddedness, it is hardly ever the case that we can determine whether the effects are due to learning, control, or a combination of the two mechanisms.

Effects of dyadic embeddedness

Gulati (1995a, 1995b) employs data on strategic alliances between business firms. Such alliances typically involve incentives for opportunism and trust problems between the partners. He finds that the probability that firms form alliances is larger if they have been previously involved in alliances with the same partner. Gulati interprets this as an indication that previous and presumably positive experiences enlarge trust among partners. Moreover, the probability that partners in alliances use equity as a formal governance mechanism and commitment device decreases with the number of previous alliances between the partners. Using other data, Gulati and Wang (2003) show that joint ventures with a longer positive relation generate more value than joint ventures with less dyadic embeddedness. This is another indication that dyadic embeddedness helps to reach more efficient solutions in the social dilemmas the partners face. More precisely, it is tempting to assume learning effects due to dyadic embeddedness as an underlying mechanism. Baker, Faulkner, and Fischer (1998) find that interorganizational ties between advertising agencies and their clients have a smaller probability of being dissolved if they have already existed for a longer period. Although these findings are interpreted in terms of learning – positive past experience increases trust, while trust enlarges the probability to stay together – a control interpretation seems likewise plausible. After all, the increased probability to stay together improves control opportunities.

Some studies on trust in inter-firm relations are noteworthy for addressing the effects of embeddedness at the dyadic level on the investment in formal arrangements such as investments in contracting. We consider investments in formal contractual arrangements as an indication for lack of trust among partners since such arrangements provide, for example, compensation for the trustor in case of untrustworthy behavior by the trustee. This can also be interpreted as that there exist *substitution effects* between contracting and embeddedness in facilitating transactions that involve trust problems. In a study on 72 subcontracting relationships, Lyons (1994) finds that the probability for arranging the relationship with a formal contract decreases with the number of years subcontractors have been trading with their most important customers. Similarly, Corts and Singh (2004) find that repeated interactions between oil companies and off-shore drillers reduce the probability that they choose fixed-price contracts to arrange the transaction. Blumberg (1997, chap. 4.2) uses the investment in formal arrangements and the extensiveness of the contract as measures for distrust in R&D-relations. He finds that both measures decrease and, thus, that trust increases with the extent to which the partners had transactions in the past. These results support the learning hypothesis that positive experiences increase trust. Blumberg actually distinguishes

trustor, i.e., given that both trustors send similar amounts, the trustee returns more or less similar amounts. Second, trustors often do not send much to trustees who do not return much.

between the effect of past transactions and transactions expected with the partner in the future, but he does not find an effect of the transactions partners expect in the future.

Batenburg, Raub, and Snijders (2003) study relations between buyers and suppliers of IT products. Their dependent variable is a combination of time and money spent in partner search, negotiating with the partner, and the extensiveness of the contract. This dependent variable represents investments in the ex-ante planning of transactions. They find that these investments decrease if the partners had transactions in the past. Furthermore, they find that the investments decrease even more if the partners already had past transactions *and* expect more transactions in the future. They do not find an effect of expected future transactions if the partners had no previous transactions. Their explanation employs two arguments. First, costly investments in ex-ante planning are less necessary if more future transactions are expected because of the sanction opportunities from subsequent transactions. This is a control effect based on the expectation of future transactions. Second, however, it is worthwhile to invest more in formal arrangements if more future transactions are expected, because these investments can be used again in subsequent transactions. This is an investment effect due to the expectation of future transactions. The driving force of this effect is that relation-specific investments associated with a focal transaction affect the incentive structure of future transactions. Combining the arguments on control and on the investment effect, it is unclear what the total effect of future transactions is. However, a negative interaction effect between past and future on ex-ante planning is indeed expected, since the investment effect will be larger in initial transactions compared to later transactions. Another explanation for such an interaction effect could be that control only plays a role if the partners have sufficient information about each other and that uncertainties about an unknown partner are simply too large to allow for reliance on control through future sanctions already in the first transaction.

Effects of network embeddedness

Buskens, Raub, and Weesie (2000) use the same data as Batenburg et al. (2003). They address the effects of network embeddedness on trust. They find that there are fewer issues addressed in the contract if the buyer and supplier are located closer to each other. An interpretation of this finding is that buyers and suppliers who are located closer to each other are probably embedded in a denser network. Although alternative explanations might be possible, this is a first indication that network embeddedness increases trust. Obviously, being located close to one another improves learning as well as control opportunities, so that it is unclear whether this effect is due to learning or control.

Gulati (1995b) argues that social networks help firms to obtain information about facilities and the abilities of potential partners. He indeed finds that alliances occur more often among partners who have more common ties with third parties. Gulati and Gargiulo (1999) is one of the studies which shows that specific network properties – in this case: centrality – are related to the likelihood of alliance formation. These findings can be interpreted as a result of learning as well as control effects of network embeddedness, because central actors potentially receive more information and they can also transmit more widely information in the network. Using data on strategic alliances in the biotechnology sector, Robinson and Stuart (2007) try to disentangle the effects of different mechanisms based on network embeddedness. Their dependent variable is equity participation, which is a measurement for mistrust, because it reduces incentive problems and increases formal control opportunities. Their important independent variables related to our study are the centrality of the trustor (“client”) and trustee (“agent” or “target”) in the network, the third parties trustor and trustee share in the network, and past alliances among the two partners. Robinson and Stuart provide strong evidence for effects of dyadic embeddedness and network embeddedness on trust by

explaining the use of informal network management mechanisms rather than equity participation for partners with repeated interactions, partners that are more central in the network, and partners that have more other partners in common. Again, however, it is impossible to distinguish clearly between learning and control effects because Robinson and Stuart use centrality measures such that the ties considered are symmetric and can be used for sending as well as receiving information. Moreover, the evidence in these network studies is based on the assumption that the network structure related to actual alliances corresponds largely with the network structure of communication among the relevant firms. If this assumption does not hold, learning and control can be the result of ties other than the alliance ties. Thus, these studies provide at best indirect evidence for the mechanisms implied in the theory, because there is no information on how actual information about behavior of the firms is spread among other firms.

The Internet economy confronts exchange partners with trust problems and nicely illustrates how actors try to solve trust problems through institutionalized information exchange that improves network embeddedness in a setting in which direct face-to-face contact is not sufficient. An important advantage of studies on the Internet economy is that for transactions at eBay and other platforms researchers know which information the buyers have about sellers. In addition, the large amount of transactions provides good opportunities for quantitative analyses. Both selling probability and selling price can be interpreted as measures for trust. First, if trust is too low, the transaction will not take place. Second, a good reputation of the seller increases the buyer's trust and the expected value of the product for the buyer, implying that she is willing to pay more. In one of the earliest studies, Kollock (1999) claims preliminary evidence that reputation scores have an effect on price. Resnick and Zeckhauser (2002) conclude from an overview of studies on eBay auctions that good reputation scores increase the likelihood that a product is sold but that there is no evidence that reputation scores affect price.

In subsequent studies, various statistical pitfalls in the analyses of Internet auctions are addressed. Resnick et al. (2006) develop a field experiment in which they use an experienced seller who sells under different identities and with different reputation scores. In this way, they keep constant many confounding factors, for example, seller's experience. Indeed, they find a large price premium related to the better reputation score. Also, in their extensive review of existing empirical work, Resnick et al. show that the price effect becomes more apparent in the more sophisticated studies. In addition, Snijders and Zijdemans (2004) show that unobserved heterogeneity can obscure the effect of reputation on price (see also Diekmann and Wyder 2002). Lucking-Reiley et al. (2007) find evidence that looking only at the net reputation score, i.e., positive minus negative evaluations is not sufficient, because negative evaluations have a much larger impact on the price than positive evaluation. Most data collected from eBay and other standard auction sites have another problem, namely, that they include information about completed deals, while there is no or at best limited information about the alternatives buyers had. This implies that the actual choice problem of buyers cannot be properly evaluated, which might lead to serious selection problems. Snijders and Weesie (2007) solve this problem by looking at a different type of site at which producers of software can offer their services for specific demands by the buyer. At the end of the auction, the buyer chooses the producer she prefers most. Snijders and Weesie (2007) again find that better reputation scores have a positive effect on price.

The evidence on Internet auctions shows that institutionalized information exchange indeed improves trust of buyers in sellers. Clearly, this can be interpreted as a learning effect: positive past information convinces buyers that a seller will act trustworthily. Still, institutionalized information exchange is likewise related to control through network embeddedness. Given that positive reputations have a price premium, buyers can damage

reputations of sellers if sellers do not perform well. However, the evidence on how negative evaluations should be weighted against positive ones and whether negative evaluations have a larger negative impact on price for sellers with good reputations than for sellers with less well-developed reputations is not so clear yet. Note that we do not address the question why people provide feedback at all. Clearly, this involves a collective good dilemma in itself since providing feedback is costly, while there is no direct benefit for oneself of providing feedback. Solving this dilemma would bring us into the literature on self-organizing institutions (see Ostrom 1990; Greif 2006; and Janssen 2006 for a specific model related to Internet auctions).

The evidence on learning and control effects on trust through social networks or more formal institutions that facilitate information exchange is still far from conclusive. While there is evidence that trust can emerge in dense social networks, it remains unclear what drives the emergence of trust. Is it learning or is control through the promise of positive and the threat of negative sanctions more important? Presumably, the empirical evidence is also limited due to scarce theoretical explanations that can guide the search for empirical evidence. Researchers have primarily focused on establishing the relationship between network embeddedness and trust considering at most one mechanism that drives this relationship. The distinction between learning and control effects of embeddedness proposed here and the fact that embeddedness facilitates learning as well as control, however, asks for an integrated approach that allows for disentangling these two mechanisms. Table 3 offers a summary of key references to empirical research on embeddedness effects.

Vignette Experiments for Disentangling Control and Learning Effects of Embeddedness

Distinguishing empirically between control and learning effects of embeddedness is a complex task. While laboratory experiments mostly allow variation in only a small number of variables, survey research often lacks the necessary control on causes and consequences. As a complement to lab experiments and survey studies, we discuss two vignette experiments in which subjects are presented with hypothetical economic transactions that involve trust problems. The subjects answer questions about their behavior related to these transactions (see, e.g., Rossi and Nock 1982 on vignette experiments). Vignette experiments are useful in providing more control over the variation of somewhat more key variables as well as over what the causes of changes in the dependent variable are. In addition, in surveys and experiments, actors are engaged in series of transactions in which opportunities for learning and control often co-occur, while in vignette experiments it is more straightforward to vary opportunities for learning and control independently. These advantages of vignette designs, however, come with two major disadvantages of vignette research. First, the choices are purely hypothetical, which implies that the choices do not have any actual consequences for the decision makers. This questions the actual incentives to choose one or the other option. Second, given that the decision situations are hypothetical, it can be a problem that the decision situation is rather artificial for the decision maker, which compromises the validity of the decisions.

In the vignette experiments, subjects have to imagine themselves in the role of buyers in economic transactions. The description of the situation makes it plausible that buyers face a trust problem by indicating that the transaction partner might have an incentive to behave opportunistically in the transaction.

TABLE 3. Key references on evidence about embeddedness effects; type of data in brackets.

Research designs and mechanisms	Two types of embeddedness	
	Dyad	Network
Experiments		
Control	Camerer and Weigelt 1988; Neral and Ochs 1992; Anderhub et al. 2002; Brandts and Figueras 2003 <i>(finitely repeated Trust Game)</i> Engle-Warnick and Slonim 2004, 2006 <i>(indefinitely as well as finitely repeated Trust Game)</i> Cochard et al. 2004 <i>(finitely repeated Investment Game)</i>	Bolton et al. 2004; Bohnet and Huck 2004 <i>(one-shot Trust Game finitely repeated within a group, while new partners obtain information about everyone's behavior in the past)</i> Barrera 2005; Buskens and Barrera 2006 <i>(finitely repeated Trust Game, trustors also obtain information about behavior of their trustee with another trustor)</i>
Learning	Camerer and Weigelt 1988; Neral and Ochs 1992; Anderhub et al. 2002; Gautschi 2002; Brandts and Figueras 2003 <i>(finitely repeated Trust Game)</i> Kollock 1994 <i>(experimental buyer-seller market)</i> Engle-Warnick and Slonim 2004, 2006 <i>(indefinitely as well as finitely repeated Trust Game)</i> Cochard et al. 2004 <i>(finitely repeated Investment Game)</i>	Bolton et al. 2004; Bohnet and Huck 2004 <i>(one-shot Trust Game finitely repeated within a group, while new partners obtain information about everyone's behavior in the past)</i> Barrera 2005; Buskens and Barrera 2006 <i>(finitely repeated Trust Game, trustors also obtain information about behavior of their trustee with another trustor)</i>
Surveys		
Control	Blumberg 1997 <i>(R&D alliances)</i> Batenburg et al. 2003 <i>(IT transactions)</i>	
Control and/or learning	Gulati 1995a, 1995b <i>(strategic alliances)</i> Gulati and Wang 2003 <i>(joint ventures)</i> Baker et al. 1998 <i>(advertising agencies and their clients)</i> Lyons 1994 <i>(subcontractors)</i> Corts and Singh 2004 <i>(off-shore drillers and oil companies)</i> Robinson and Stuart 2007 <i>(strategic alliances)</i>	Buskens et al. 2000 <i>(IT transactions, geographical proximity as density measure)</i> Gulati 1995b; Gulati and Gargiulo 1999 <i>(R&D alliances, third-party relations, centrality)</i> Robinson and Stuart 2007 <i>(strategic alliances, centrality, proximity)</i> Resnick and Zeckhauser 2002; Diekmann and Wyder 2002; Snijders and Zijdemann 2004; Resnick et al. 2006; Lucking-Reiley et al. 2007; Snijders and Weesie 2007 <i>(Internet auctions)</i>
Learning	Blumberg 1997 <i>(R&D alliances)</i> Batenburg et al. 2003 <i>(IT transactions)</i>	

In the first experiment, purchase managers of Dutch companies are asked to answer questions about hypothetical transactions with suppliers (see Rooks et al. 2000). The description of the transactions comprises information about transaction characteristics such as price and specific investments of the buyer associated with the transaction, but also about the relationship of the buyer with the supplier. Five variables are varied that are related to embeddedness:

- The extent to which the buyer did business with the same supplier in the past (dyadic learning).
- The extent to which the buyer expects to do business with the supplier in the future (dyadic control).
- The extent to which the buyer and supplier have common business partners (network embeddedness that provides opportunities for learning as well as control).
- The availability of alternative suppliers for the buyer (network control).

The dependent variable is lack of trust of the buyer, measured by the extent to which she wants to invest in safeguards (e.g., contractual agreements) before the transaction takes place. Results reveal a strong effect of embeddedness on trust due to learning within a dyadic relation. Positive past experiences reduce the investment in safeguards. Although there is no main effect of expected interactions in the future, there is indeed a negative interaction effect of past transactions and expected future transactions, indicating that the occurrence of control is contingent on some previous learning opportunities. This finding is nicely in line with the results of the survey on IT transactions of Batenburg et al. (2003) discussed above, notably employing a very different research design. Concerning third-party effects, results show that knowing other business partners of the supplier increases trust. It is unclear whether this effect is due to learning or control, since these third parties can be used to obtain information on previous behavior of the supplier, but they also can be informed on behavior of the supplier in the focal transaction, thus extending control opportunities for the buyer. There is indeed a negative effect of the availability of alternative trading partners on investments of the buyer in safeguards for the transaction. This supports the interpretation that purchase managers realize that alternative suppliers provide them with sanction opportunities, implying that the supplier is less likely to act untrustworthy if he has more competitors.

In another vignette experiment, students are asked to compare situations for buying a used car (see Buskens and Weesie 2000b for more details). Students are offered pairs of vignettes describing such a transaction and they are asked which one they preferred. Five embeddedness variables are varied at the vignettes:

- Whether the buyer has bought a car from the dealer before and was satisfied, or did never buy a car from the dealer (dyadic learning).
- Whether or not the buyer expects to move to the other side of the country soon (dyadic control). The probability that the buyer has future transactions with the dealer is smaller if the buyer moves. Hence, control is more difficult for a buyer if she moves. Theoretically, the effect of expected future transactions is based on the sanctions of the buyer *anticipated by the dealer*. Therefore, strictly speaking, expected future transactions can be expected to affect the behavior of buyer and dealer only if the dealer is informed about the buyer's plans to move.
- Whether the dealer is or is not well-known in the neighborhood of the buyer (network density). Again, learning as well as control of a well-known garage through the network of customers can be more effective than learning about or control of a garage that is not well-known.

- Whether or not the buyer has information from friends about transactions of these friends with the garage (network learning). The focus is on the difference between *no* information and *positive* information.
- Whether or not both the buyer and the dealer are members of the same sports team (network control). This measures network control because the number of acquaintances the buyer and dealer have in common is expected to be larger if the buyer and dealer are members of the same sports team. Common membership provides the buyer with possibilities of controlling the dealer through positive or negative reputational sanctions both in his business and as a team member. An advantage of this operationalization is that the theoretical assumption of “common knowledge about the network” is unlikely to be violated because the buyer and the dealer both know that they are members of the sports team. DiMaggio and Louch (1998) demonstrate that many buyers prefer a relative as a dealer for a used car rather than a dealer with whom they have no social relationship.

Results show that all five embeddedness variables have positive effects on the likelihood that subjects prefer a vignette that includes the respective type of embeddedness over the one in which that type of embeddedness is not available. The strongest effects seem again to be those of dyadic and network learning variables. Positive information clearly enhances trust. Dyadic control, density, and network control have likewise positive effects on trust, which implies that control is important at the dyadic as well as at the network level. The evidence for a control mechanism is somewhat problematic since these variables are subject to alternative explanations. There might be other disadvantages for buying a car just before you move, for example, because you need to find another garage if there is any problem with the car in the future. Although our arguments and results for network control are in accordance with DiMaggio and Louch (1998), we cannot exclude that actors prefer to trust well-known others over unknown others for other reasons than the control reasons advocated here.

Summarizing the evidence from studies employing different research designs, we have quite unambiguous support for hypotheses on learning effects at the dyadic and network level. Also, hypotheses on effects of control opportunities at the dyadic level are quite consistently supported, as can be seen from end-game effects in finitely repeated Trust Games, surveys on transactions as well as the vignette experiments. Network control is less well studied and the evidence is more ambiguous. We would expect that these results might generalize to other kinds of social dilemmas, but we do not know about systematic studies that have compared embeddedness effects for other social dilemmas, including experimental as well as survey studies and distinguishing between different types of embeddedness effects.

CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

We have provided a survey of rational choice research on social dilemmas by focusing on how game theory can be used to model social dilemmas, how testable hypotheses can be generated from game-theoretic models, and what empirical evidence tells us about those hypotheses. Trust problems have been our paradigm case of a social dilemma. In terms of the strategies for refining the model of atomized interactions on perfect markets of rational and selfish actors with full information, we focused on models that retain the rationality assumption. In fact, game-theoretic models often employ particularly strong rationality assumptions. We briefly considered how relaxing selfishness assumptions by including other-regarding preferences can help in accounting for behavior in social dilemmas and specifically in dilemmas that are isolated encounters. Our main focus, however, has been on effects of social embeddedness. We concentrated on game-theoretic models that allow for an analysis of

how embeddedness affects behavior in social dilemmas. Hence, the bulk of the models surveyed in this chapter relax the assumption of atomized interactions and often also the assumption of full information. We have stressed that game-theoretic models allow to systematically distinguish different kinds of embeddedness and to also distinguish different mechanisms such as control and learning through which behavior in social dilemmas depends on embeddedness.

From the empirical end, we have stressed the need for research designs that allow to discriminate between different types of embeddedness as well as to disentangle control and learning effects. We also argued for using complementary research designs such as experiments, surveys, vignette studies, etc. as a strategy for establishing the robustness of empirical findings (see Levitt and List 2007 for a thorough discussion of this issue). Our overview of studies shows that there is quite some empirical evidence for embeddedness effects. When research designs are employed that do allow for disentangling different kinds of embeddedness effects and mechanisms through which embeddedness works, hypotheses based on game-theoretic models often succeed in predicting the signs of coefficients (see, e.g., Grofman 1993 and Green and Shapiro 1994 for related discussion on the merits and problems of qualitative predictions on changes “at the margin” using comparative statics versus quantitative point predictions from rational choice models). Nevertheless, there is clearly much room for improvement of the predictions of game-theoretic models on behavior in social dilemmas. Roughly speaking, the overall impression is that assuming game-theoretic rationality as well as selfish actors (“utility = own money”) cannot account for quite some non-opportunistic behavior in social dilemmas that are isolated encounters, while it also often predicts “too much” cooperative behavior in repeated social dilemmas (see, e.g., Bolton and Ockenfels 2006 for a similar point in the context of research on reputation systems in the Internet economy). Developing game-theoretic models on the interplay of social embeddedness and other-regarding preferences may be useful in this respect (see, e.g., Gintis 2000, chap. 11 for related arguments).

With respect to research on social embeddedness, theoretical as well as empirical work reviewed in this chapter assumed embeddedness characteristics as exogenously given. Using a notion that has become popular, embeddedness can be conceived as social capital of actors (e.g., Coleman 1990). We have focused on the returns on social capital: embeddedness allows for overcoming Pareto-suboptimal outcomes in social dilemmas. What we have neglected are actors’ investments in their social capital (see, e.g., Flap 2004 for the distinction between returns on and investments in social capital). However, given the returns on embeddedness in social dilemmas, actors do have an incentive to invest in their embeddedness by strategically establishing, maintaining, or deleting ties to others, including search for potential interaction partners. One would thus like to endogenize embeddedness characteristics. Research on strategic network formation based on game-theoretic models is rather novel but meanwhile rapidly developing (see the edited volume Dutta and Jackson 2003 as well as Goyal’s 2007 textbook). Such work on returns on and investments in social capital can likewise benefit from the development of “actor-driven” statistical models for the dynamics (“co-evolution”) of networks and behavior (Snijders 2001 and Snijders’ chapter in this Handbook).

How embeddedness can contribute to trust and cooperation has been a core topic of our chapter. We thus highlighted the beneficial effects of embeddedness for the actors involved in a social dilemma. It has already been mentioned that such beneficial effects for the actors directly involved can have negative effects for others. From the perspective of third parties or from a societal perspective, undermining rather than fostering cooperation is often the aim. It should be noted, too, that embeddedness can also have adverse effects for the actors who are themselves directly involved in social dilemmas. Focusing on learning and information diffusion rather than game-theoretic rationality as a driving force of behavior, Burt and Knez

(1995) have shown that dense networks can amplify trust as well as distrust. The core argument is that due to the homogeneity of opinions in a dense network, actors become convinced about some information because they receive the information disproportionately often. Co-ethnics may be able to solve trust problems in economic exchange by transacting with each other, but this may lead to entrapment and missing opportunities from outside networks (e.g., Portes 1998). Flache (2002) offers a game-theoretic model of how informal social ties between the members of a team can undermine cooperation of the members of the team since they have to trade-off the benefits of sanctioning team members who do not cooperate against the costs of deteriorating informal social ties through negative sanctions. While there is quite some empirical research on adverse effects of embeddedness, more systematic theoretical modeling of such effects is needed.

With respect to theoretical modeling, relaxing strong game-theoretic rationality assumptions or showing that and when equilibrium behavior in accordance with such assumptions is a – possibly long-term – result of bounded rationality and evolutionary or learning processes (e.g., Fudenberg and Levine 1998; Gintis 2000) may have useful implications for research on social dilemmas, too. We would like to conclude, though, with a more specific suggestion. We have seen that, in principle, games with incomplete information are a tool for analyzing embeddedness effects in social dilemmas through control and learning in an integrated way. However, as we have also seen, it is difficult to strike a fruitful balance of analytic tractability and realistic assumptions about what information actors have and how they use relevant sources of information. On the one hand, models with more realistic informational assumptions are often difficult to analyze. On the other hand, knowledge about what realistic informational assumptions would be is limited, because most empirical research has not succeeded in clearly disentangling the effects of learning and control mechanisms. Disentangling the effects could provide some evidence on the relative importance of these mechanisms and evidence about changes in the importance of different effects related to different circumstances.

To overcome these limitations, we propose a two-step empirical and theoretical approach. More experimental research is necessary to obtain better insights in the relative importance of the different mechanisms. E.g., experiments should be designed such that subjects are involved in abstract Trust Games embedded in a social context that allows for communication among trustors. The experiments should explicitly provide insights in how subjects use information they obtain from other subjects in the network and whether or not they try to sanction by informing other trustors in the network. In this way, the experiments enable the development of new models built on assumptions, for example, about information exchange that have an empirical basis rather than on assumptions chosen exclusively on the basis of introspection of researchers and mathematical tractability. Moreover, the experiments can be used to obtain initial insights in circumstances that affect the importance of control versus learning. Results of such experiments can inspire new theoretical models on the relative effects of learning versus control. Based on these models, survey designs can then be developed that allow for variations in learning and control variables such that the predicted effects can be distinguished.

REFERENCES

- Anderhub, Vital, Dirk Engelmann, and Werner Güth. 2002. “An Experimental Study of the Repeated Trust Game with Incomplete Information.” *Journal of Economic Behavior and Organization* 48:197–216.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

- Bacharach, Michael and Diego Gambetta. 2001. "Trust in Signs." Pp. 148–84 in *Trust in Society*, ed. Karen S. Cook. New York: Russell Sage.
- Baker, Wayne E., Robert R. Faulkner, and Gene A. Fisher. 1998. "Hazards of the Market: The Continuity and Dissolution of Interorganizational Market Relationships." *American Sociological Review* 63:147–77.
- Barber, Bernard. 1983. *The Logic and Limits of Trust*. New Brunswick, NJ: Rutgers University.
- Barrera, Davide. 2005. *Trust in Embedded Settings*. Veenendaal: Universal Press.
- Barrera, Davide and Vincent Buskens. 2006. *Third-Party Effects in an Embedded Investment Game*. ISCORE paper 226, Utrecht University.
- Batenburg, Ronald S., Werner Raub, and Chris Snijders. 2003. "Contacts and Contracts: Temporal Embeddedness and the Contractual Behavior of Firms." *Research in the Sociology of Organizations* 20:135–88.
- Bendor, Jonathan and Dilip Mookherjee. 1987. "Institutional Structure and the Logic of Ongoing Collective Action." *American Political Science Review* 81:129–54.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10:122–42.
- Binmore, Ken. 1998. *Game Theory and the Social Contract, Volume 2: Just Playing*. Cambridge, MA: MIT Press.
- Blau, Peter M. [1964] 1996. *Exchange and Power in Social Life*. New Brunswick, NJ: Transaction Publishers.
- Blumberg, Boris F. 1997. *Das Management von Technologiekoooperationen: Partnersuche und Verhandlungen mit dem Partner aus empirisch-theoretischer Perspektive*. Amsterdam: Thesis Publishers.
- Bohnet, Iris. 1997. *Kooperation und Kommunikation*. Tübingen: Mohr.
- Bohnet, Iris and Steffen Huck. 2004. "Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change." *American Economic Review (Papers and Proceedings)* 94:362–66.
- Bolton, Gary E., Elena Katok, and Axel Ockenfels. 2004. "How Effective Are Online Reputation Mechanisms? An Experimental Study." *Management Science* 50:1587–602.
- Bolton, Gary E. and Axel Ockenfels. 2006. *The Limits of Trust in Economic Transactions. Investigations of Perfect Reputation Systems*. Working paper.
- Bower, Anthony, Steven Garber, and Joel C. Watson. 1997. "Learning about a Population of Agents and the Evolution of Trust and Cooperation." *International Journal of Industrial Organization* 15:165–90.
- Brandts, Jordi and Neus Figueras. 2003. "An Exploration of Reputation Formation in Experimental Games." *Journal of Economic Behavior and Organization* 50:89–115.
- Burt, Ronald S. and Marc Knez. 1995. "Kinds of Third-Party Effects on Trust." *Rationality and Society* 7:255–92.
- Buskens, Vincent. 2002. *Social Networks and Trust*. Boston, MA: Kluwer.
- Buskens, Vincent. 2003. "Trust in Triads: Effect of Exit, Control, and Learning." *Games and Economic Behavior* 42:235–52.
- Buskens, Vincent and Werner Raub. 2002. "Embedded Trust: Control and Learning." *Advances in Group Processes* 19:167–202.
- Buskens, Vincent, Werner Raub, and Jeroen Weesie. 2000. "Networks and Contracting in Information Technology Transactions." Pp. 77–81 in *The Management of Durable Relations: Theoretical and Empirical Models for Organizations and Households*, ed. Werner Raub and Jeroen Weesie. Amsterdam: Thela Thesis.
- Buskens, Vincent and Jeroen Weesie. 2000a. "Cooperation via Networks." *Analyse & Kritik* 22:44–74.

- Buskens, Vincent and Jeroen Weesie. 2000b. "An Experiment on the Effects of Embeddedness in Trust Situations: Buying a Used Car." *Rationality and Society* 12:227–53.
- Camerer, Colin F. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. New York: Russell Sage.
- Camerer, Colin F. and Ernst Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists. Pp. 55–95 in *Foundations of Human Sociality: Experimental and Ethnographic Evidence from 15 Small-Scale Societies*, ed. Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. Oxford: Oxford University Press.
- Camerer, Colin F. and Keith Weigelt. 1988. "Experimental Tests of a Sequential Equilibrium Reputation Model." *Econometrica* 56:1–36.
- Cochard, Francois, Phu Nguyen Van, and Marc Willinger. 2004. "Trusting Behavior in a Repeated Investment Game." *Journal of Economic Behavior and Organization* 55:31–44.
- Coleman, James S. 1964. "Collective Decisions." *Sociological Inquiry* 34:166–81.
- Coleman, James S. 1987. "Free Riders and Zealots." Pp. 59–82 in *Social Exchange Theory*, ed. Karen S. Cook. Newbury Park, CA: Sage.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.
- Colman, Andrew M. 1982. *Game Theory and Experimental Games: The Study of Strategic Interactions*. Oxford: Pergamon Press.
- Cook, Karen S. and Robin M. Cooper. 2003. "Experimental Studies of Cooperation, Trust, and Social Exchange." Pp. 209–44 in *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, ed. Elinor Ostrom and James Walker. New York: Russell Sage.
- Corts, Kenneth S. and Jasjit Singh. 2004. "The Effect of Repeated Interaction on Contract Choice: Evidence from Offshore Drilling." *Journal of Law, Economics and Organization* 20:230–60.
- Dasgupta, Partha. 1988. "Trust as a Commodity." Pp. 49–72 in *Trust: Making and Breaking Cooperative Relations*, ed. Diego Gambetta. Oxford: Blackwell.
- Dawes, Robert M. 1980. "Social Dilemmas." *Annual Review of Psychology* 31:169–93.
- Diekmann, Andreas. 2004. "The Power of Reciprocity." *Journal of Conflict Resolution* 48:487–505.
- Diekmann, Andreas and David Wyder 2002. "Vertrauen und Reputationseffekte bei Internet-Auktionen." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54:674–93.
- DiMaggio, Paul and Hugh Louch. 1998. "Socially Embedded Consumer Transactions: For What Kinds of Purchases Do People Most Often Use Networks?" *American Sociological Review* 63:619–37.
- Durkheim, Emiel. [1893] 1973. *De la Division du Travail Social* (9th edition). Paris: PUF.
- Dutta, Bhaskar. and Matthew O. Jackson, eds. 2003. *Networks and Groups. Models of Strategic Formation*. Berlin: Springer.
- Engle-Warnick, Jim and Robert L. Slonim. 2004. "The Evolution of Strategies in a Repeated Trust Game." *Journal of Economic Behavior and Organization* 55:553–73.
- Engle-Warnick, Jim and Robert L. Slonim. 2006. "Learning to Trust in Indefinitely Repeated Games." *Games and Economic Behavior* 54:95–114.
- Fehr, Ernst and Herbert Gintis. 2007. "Human Motivation and Social Cooperation: Experimental and Analytical Foundations." *Annual Review of Sociology* 33:43–64.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108:437–59.

- Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114:817–68.
- Fehr, Ernst and Klaus M. Schmidt. 2005. *The Rhetoric of Inequity Aversion – A Reply*. Working paper.
- Fehr, Ernst and Klaus M. Schmidt. 2006. "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories." Pp 615–91 in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. Serge-Christophe Kolm and Jean Mercier Ythier. Amsterdam: Elsevier.
- Flache, Andreas 2002. "The Rational Weakness of Strong Ties." *Journal of Mathematical Sociology* 26:189–216.
- Flache, Andreas, and Michael W. Macy. 2002. "Stochastic Collusion and the Power Law of Learning: A General Reinforcement Learning Model of Cooperation." *Journal of Conflict Resolution* 46:629–53.
- Flap, Henk. 2004. "Creation and Returns of Social Capital." Pp. 3–23 in *Creation and Returns of Social Capital*, ed. Henk Flap and Beate Völker. London: Routledge.
- Friedman, James W. 1971. "A Non-Cooperative Equilibrium for Supergames." *Review of Economic Studies* 38:1–12.
- Friedman, James W. 1990. *Game Theory with Applications to Economics* (2nd edition). New York: Oxford University Press
- Fudenberg, Drew and David K. Levine. 1998. *The Theory of Learning in Games*. Cambridge, MA: MIT Press.
- Fudenberg, Drew and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54:533–54.
- Gächter, Simon. 2008. Chapter in this Handbook.
- Gächter, Simon and Armin Falk. 2002. "Reputation and Reciprocity: Consequences for the Labour Relation." *Scandinavian Journal of Economics* 104:1–26.
- Gautschi, Thomas. 2000. "History Effects in Social Dilemma Situations." *Rationality and Society* 12:131–62.
- Gibbons, Robert. 2001. "Trust in Social Structures: Hobbes and Coase Meet Repeated Games." Pp. 332–53 in *Trust in Society*, ed. Karen S. Cook. New York: Russell Sage.
- Gintis, Herbert. 2000. *Game Theory Evolving*. Princeton, NJ: Princeton University Press.
- Gintis, Herbert. 2007. "A Framework for the Unification of the Behavioral Sciences." *Behavioral and Brain Sciences* 30:1–61.
- Goldthorpe, John H. 2000. *On Sociology. Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.
- Gouldner, Alvin W. 1960. "The Norm of Reciprocity." *American Sociological Review* 25:161–78.
- Goyal, Sanjeev. 2007. *Connections. An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press.
- Granovetter, Mark S. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91:481–510.
- Green, Donald P. and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.
- Green, Edward J. and Robert H. Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica* 52:87–100.
- Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49:857–82.
- Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.
- Greif, Avner. 2008. Chapter in this Handbook.

- Grofman, Bernard. 1993. "Is Turnout the Paradox That Ate Rational Choice Theory?" Pp. 93–103 in *Information, Participation, and Choice*, ed. Bernard Grofman. Ann Arbor, MI: University of Michigan Press.
- Gulati, Ranjay. 1995a. "Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances." *Academy of Management Journal* 38:85–112.
- Gulati, Ranjay. 1995b. "Social Structure and Alliance Formation Patterns: A Longitudinal Study." *Administrative Science Quarterly* 40:619–52.
- Gulati, Ranjay and Martin Gargiulo. 1999. "Where Do Interorganizational Networks Come From?" *American Journal of Sociology* 104:1439–93.
- Gulati, Ranjay and Lihua Wang. 2003. "Size of the Pie and Share of the Pie: Implications of Structural Embeddedness for Value Creation and Value Appropriation in Joint Ventures." *Research in the Sociology of Organizations* 20:209–42.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3:367–88.
- Hardin, Russell. 2001. "Conceptions and Explanations of Trust." Pp. 3–39 in *Trust in Society*, ed. Karen S. Cook. New York: Russell Sage.
- Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage.
- Harsanyi, John C. 1967/68. "Games with Incomplete Information Played by 'Bayesian' Players I-III." *Management Science* 14:159–82, 320–34, 486–502.
- Harsanyi, John C. 1975. *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht: Reidel.
- Harsanyi, John C. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.
- Heckathorn, Douglas D. 2008. Chapter in this Handbook.
- Van der Heijden, Eline C.M., Jan H.M. Nelissen, Jan J.M. Potters, and Harrie A.A. Verbon. 2001. "Simple and Complex Gift Exchange in the Laboratory." *Economic Inquiry* 39:280–97
- Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty. Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.
- Hirshleifer, David and Eric Rasmusen. 1989. "Cooperation in a Repeated Prisoner's Dilemma with Ostracism." *Journal of Economic Behavior and Organization* 12:87–106.
- Hobbes, Thomas. [1651] 1991. *Leviathan*. Cambridge: Cambridge University Press.
- Janssen, Marco. 2006. "Evolution of Cooperation when Feedback to Reputation Scores Is Voluntary." *Journal of Artificial Societies and Social Simulation* 9: <jass.soc.surrey.ac.uk/9/1/17.html>.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76:728–41.
- Klein, Daniel B. 1997. *Reputation: Studies in the Voluntary Elicitation of Good Conduct*. Ann Arbor, MI: University of Michigan Press.
- Kollock, Peter. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." *American Journal of Sociology* 100:313–45.
- Kollock, Peter. 1998. "Social Dilemmas: The Anatomy of Cooperation." *Annual Review of Sociology* 24:183–214.
- Kollock, Peter 1999. "The Production of Trust in Online Markets." *Advances in Group Processes* 16:99–123.
- Kreps, David M. 1990a. "Corporate Culture and Economic Theory." Pp. 90–143 in *Perspectives on Positive Political Economy*, ed. James E. Alt and Kenneth A. Shepsle. Cambridge: Cambridge University Press.
- Kreps, David M. 1990b. *Game Theory and Economic Modelling*. Oxford: Clarendon Press.
- Kreps, David M. and Robert Wilson 1982. "Sequential Equilibria," *Econometrica* 50:863–94.

- Ledyard, John O. 1995. "Public Goods: A Survey of Experimental Research." Pp 111–94 in *The Handbook of Experimental Economics*, ed. John. H. Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press.
- Levitt, Steven D. and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preference Reveal About the Real World?" *Journal of Economic Perspectives* 21:153–74.
- Liebrand, Wim B.G. 1983. "A Classification of Social Dilemma Games." *Simulation and Games* 14:123–38.
- Lorenz, Edward H. 1988. "Neither Friends Nor Strangers: Informal Networks of Subcontracting in French Industry." Pp. 94-107 in *Trust: Making and Breaking Cooperative Relations*, ed. Diego Gambetta. Oxford: Blackwell.
- Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves 2007. "Pennies from eBay: The Determinants of Price in Online Auctions." *Journal of Industrial Economics* 55:223–33.
- Lyons, Bruce R. 1994. "Contracts and Specific Investments: An Empirical Test of Transaction Cost Theory." *Journal of Economics and Management Strategy* 3:257–78.
- Macy, Michael W. and Andreas Flache. 1995. "Beyond Rationality in Models of Choice." *Annual Review of Sociology* 21:73–91.
- Macy, Michael W. and Andreas Flache. 2002. "Learning Dynamics in Social Dilemmas." *Proceedings of the National Academy of Sciences U.S.A.* 99:7229–36.
- Merton, Robert K. 1973. *The Sociology of Science*, Chicago, IL: University of Chicago Press.
- Milgrom, Paul, Douglas C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchants. *Economics and Politics* 2:1–23.
- Nash, John F. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54:286–95.
- Neral, John and Jack Ochs. 1992. "The Sequential Equilibrium Theory of Reputation Building: A Further Test." *Econometrica* 60:1151–69.
- Ortmann, Andreas, John Fitzgerald, and Carl Boeing. 2000. "Trust, Reciprocity, and Social History: A Re-examination." *Experimental Economics* 3:81–100.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom, Elinor. 2003. "Toward a Behavioral Theory Linking Trust, Reciprocity, and Reputation. Pp. 19–79 in *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, ed. Elinor Ostrom and James Walker. New York: Russell Sage.
- Parsons, Talcott. 1937. *The Structure of Social Action*. New York: Free Press.
- Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24:1–24.
- Rapoport, Anatol. 1974. "Prisoner's Dilemma - Recollections and Observations." Pp. 18–34 in *Game Theory as a Theory of Conflict Resolution*, ed. Anatol Rapoport. Dordrecht: Reidel.
- Rasmusen, Eric. 2007. *Games and Information: An Introduction to Game Theory* (4th edition). Oxford: Blackwell.
- Raub, Werner. 2004. "Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling." *Rationality and Society* 16:319–66.
- Raub, Werner and Jeroen Weesie. 1990. "Reputation and Efficiency in Social Interactions: An Example of Network Effects." *American Journal of Sociology* 96:626–54.
- Resnick, Paul and Richard Zeckhauser. 2002. "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." *Advances in Applied Microeconomics* 11:127–57.
- Resnick, Paul, Zeckhauser, Richard, Swanson, John, and Kate Lockwood. 2006. "The Value of Reputation on eBay: A Controlled Experiment." *Experimental Economics* 9:79–101.

- Van de Rijt, Arnout and Michael W. Macy. 2006. "Social Dilemmas: Neither More Nor Less." Working Paper.
- Robinson, David T., and Toby E. Stuart. 2007. "Network Effects in the Governance of Strategic Alliances." *Journal of Law, Economics and Organization* 23:242–73.
- Rooks, Gerrit, Werner Raub, Robert Selten, and Frits Tazelaar. 2000. "Cooperation between Buyer and Supplier: Effects of Social Embeddedness on Negotiation Effort." *Acta Sociologica* 43:123–37.
- Rossi, Peter H. and Steven L. Nock. eds. 1982. *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills, CA: Sage.
- Sally, David. 1995. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society* 7:58–92.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. London: Oxford University Press.
- Schramm, Matthias and Markus Taube. 2003. "Evolution and Institutional Foundation of the Hawala Financial System." *International Review of Financial Analysis* 12:405–20.
- Schübler, Rudolf A. 1989. "Exit Threats and Cooperation under Anonymity." *Journal of Conflict Resolution* 33:728–49.
- Selten, Reinhard. 1965. "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragerträgeit." *Zeitschrift für die gesamte Staatswissenschaft* 121:301–24, 667–89.
- Shaked, Avner. 2006. *On the Explanatory Value of Inequity Aversion Theory*. Working paper.
- Snijders, Chris. 1996. *Trust and Commitments*. Amsterdam: Thesis Publishers.
- Snijders, Chris and Gideon Keren 1999. "Determinants of Trust." Pp. 355–85 in *Games and Human Behavior*, ed. David V. Budescu, Ido Erev, and Rami Zwick. Mahwah, NJ: Lawrence Erlbaum.
- Snijders, Chris and Gideon Keren 2001. "Do You Trust? Whom Do You Trust? When Do You Trust?" *Advances in Group Processes* 18:129–60.
- Snijders, Chris and Jeroen Weesie. 2007. *Trust and Reputation in an Online Programmer's Market*. Working paper.
- Snijders, Chris and Richard Zijdemán. 2004. "Reputation and Internet Auctions: eBay and beyond." *Analyse & Kritik* 26:158–84.
- Snijders, Tom A.B. 2001. "The Statistical Evaluation of Social Network Dynamics." *Sociological Methodology* 29:361–95.
- Snijders, Tom A.B. 2008. Chapter in this Handbook.
- Taylor, Michael. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press (revised edition of *Anarchy and Cooperation*. London: Wiley 1976).
- Uzzi, Brian. 1996. "The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect." *American Sociological Review* 61:674–98.
- Vanberg, Viktor J. and Roger D. Congleton. 1992. "Rationality, Morality, and Exit." *American Political Science Review* 86:418–31.
- Voss, Thomas. 1982. "Rational Actors and Social Institutions: The Case of the Organic Emergence of Norms." Pp. 76–100 in *Theoretical Models and Empirical Analyses. Contributions to the Explanation of Individual Actions and Collective Phenomena*, ed. Werner Raub. Utrecht: ESP.
- Weber, Max. [1921] 1976. *Wirtschaft und Gesellschaft* (5th edition). Tübingen: Mohr.
- Weber, Max. 1947. *The Theory of Social and Economic Organization*. New York: Free Press.
- Wechsberg, Joseph. 1966. *The Merchant Bankers*. New York: Bedminster Press.
- Weesie, Jeroen and Werner Raub 1996. "Private Ordering: A Comparative Institutional Analysis of Hostage Games," *Journal of Mathematical Sociology* 21:201–40.
- Weesie, Jeroen, Vincent Buskens, and Werner Raub. 1998. "The Management of Trust Relations via Institutional and Structural Embeddedness." Pp. 113–38 in *The Problem of*

- Solidarity: Theories and Models*, ed. Patrick Doreian and Thomas Fararo. Amsterdam: Gordon and Breach.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Williamson, Oliver E. 1996. *The Mechanisms of Governance*. New York: Oxford University Press.
- Wittek, Rafael, Tom A.B. Snijders, and Victor Nee. 2008. Introduction in this Handbook.
- Yamagishi, Toshio and Midori Yamagishi. 1994. "Trust and Commitment in the United States and Japan." *Motivation and Emotion* 18:129–66.