

The economics of cyber trust between cyber partners

Jonathan Cave
University of Warwick

While the Office of Science and Technology commissioned this review, the views are those of the authors, are independent of Government and do not constitute Government policy.

1 INTRODUCTION

This paper considers some aspects of the intersection of trust and economics through the application of economic tools to the analysis of trust and the delineation of trust issues in the analysis of economics. As far as possible, the development of the argument will be related to trust in electronic contexts, though the former deals with more abstract issues that are necessarily less directly influenced by the specific features of electronic interactions than the latter.

This paper is divided into two parts. The first part (Sections 2 and 3) applies game theoretic tools to specific aspects of trust, while Section 4 presents a brief survey of trust in industrial organization economics (because this is the most appropriate context for cyber trust, and highlights trust considerations relating to, for example, e-commerce) with just enough model development to indicate a 'road map' for further development.

Section 2 develops some simple, essentially static game theoretic models of specific aspects of trust to analyze their equilibrium and efficiency outcomes. Section 3 examines the evolution of trust as a 'convention'. The game-theoretic analysis points up certain abstract implications that transcend the directly economic setting, but are less precise and less faithful to institutional detail. The supporting models shed light on the potential tension between efficiency and equilibrium, both in terms of the prevalence of trust behaviour and the network of relationships to which the need to trust others gives rise.

The models also identify conditions under which different levels of trust may 'prevail' (be widespread) and conditions under which a diversity of behaviour is likely. One consequence is the existence of 'catastrophes' – discontinuous jumps in the level of trust in response to small changes in underlying conditions. This can give rise to cyclic behaviour. Another use of the models is identifying low-cost ways for policy to exploit the evolutionary nature of trust in order to promote efficiency.

The purpose of the summary of more conventional economic analysis in Section 4 is two-fold: to develop implications of existing economic models for trust and, conversely, to show how trust considerations have modified analysis of industrial structure, conduct and performance. In addition, this summary identifies places where game-theoretic analysis can usefully contribute to more practical policy problems. Section 5 concludes, developing policy implications and suggestions for further work.

1.1 Methodology

Several methodological points should be made at the outset. The game-theoretic models in Sections 2 and 3 seem more abstract than the economic models in Section 4 – but both approaches are theoretical and, in all cases, alternative models and conclusions are possible. None of the results should be taken as assertions about the way the world is or should be. Rather, the analysis starts from plausible representations of aspects of trust and derives conclusions using standard tools – the conclusions are only as empirically valid as the model's assumptions.²

The language of theory is necessarily condensed and may strike many readers as opaque, so details³ of the models, derivations and computations have for the most part been placed in an appendix. The utility ultimately lies in a positive contribution – *if* we conceptualize trust in a given way, certain implications follow. The material in the appendix is intended to indicate how the argumentation runs, so that its strengths and weaknesses and potential for development in more specific situations can be assessed.

The game theoretic approach starts from specific stylized aspects of trust:

- one's own level of trust should match one's environment;
- opportunistic or criminal behaviour and carelessness or suspicion have different impacts on trust;
- trust creates linkage and link formation requires trust;
- the value of a trusted relationship may be affected by the partner's other trusted relationships;
- costly trust-enhancing activities may offer external benefits;
- individuals may be able to do similar things in high- and low-trust environments;
- there are unobserved individual differences in the subjective importance of trust;
- 'network externalities' (remote effects of trust linkages) can affect the dynamics of trusting behaviour; and
- trust can be viewed as a societal convention or norm.

Such models can illuminate such abstract features as:

- multiple stable outcomes;
- the dynamic evolution of trust (resilience, jumps, cycles, etc.);
- connections between trust behaviour and the social fabric;
- the tension between stability and efficiency; and
- the scope for policy to exploit evolution and encourage efficient trust.

The economic analysis of industrial organization addresses the relations among economic actors and the degree to which competition and its efficiency consequences are affected by institutional, technological and informational factors. This offers ample scope for 'trust metaphors'. In contrast to abstract game-theoretic treatments of trust these tend to focus on concrete economic interactions, for example, some form of contract and/or market. The economic analysis sheds light on how trust affects the operation of these institutions and on how economic considerations in turn affect the level of trust. Specifically, industrial economics considers:

- reputations among economic agents that are linked,
 - horizontally – as competitors in markets;
 - vertically – as buyers and sellers of goods and services;
 - in networks – as producers and users of complimentary goods and services.
- standards relating to trust;
- the economics of security and security-enhancing technologies
- assurance and market mechanisms for signalling and providing it;
- liability and its effect on the efficiency and equity of precaution and risk-bearing.

The analysis is tied to specific market features and thus provides more concrete ways to take account of trust in economic policy and of economic factors (especially profit) in policy relating to trust.

The mapping between issues and models is not one-to-one: many aspects of cyber trust are not addressed by the models and many phenomena with which the models are concerned (particularly in Section 3.4) have only an indirect connection to trust, or the specific implications of the Internet. However, some connections are stronger than may at first appear, and we attempt to draw out the interpretation of models in terms that relate to trust. This indirect approach reflects the inherent complexity of trust and attempts to balance tractability with conceptual relevance.

1.2 Trust and Trusting

Trust is a matter of expectation – extrapolations to other times and contingencies. A trusting individual has some view of what might happen, how likely the various possibilities are and the impact of current choices. These assessments may not follow expected utility theory: likelihoods may be more subjective than objective, outcome and likelihood may not be separable and the contingent future may be described in deliberately 'fuzzy' language. Within the expected utility framework, individuals may ignore the variance of outcomes. Finally, trust may involve different degrees of consequentialism – expectations may be bound up with process as well as outcomes.

For example, an online customer may trust a transaction without distinguishing reliability of merchants, payment and/or delivery services and legal mechanisms that provide compensation in the event of loss. Others may evaluate a purchase quite differently, being reluctant to disclose payment details to some agents (but not all) even while maintaining the same beliefs about the likelihood of different outcomes.

Trust has proven a difficult concept for economics to clarify (see Hollis 1998), but its influence is widely acknowledged. Much analysis in the economic literature follows one or more dichotomies.⁴ Several considerations follow from these distinctions.

The participation decision (if alternatives are available) is essentially to trust that the game is as described, that the other players will behave as expected, etc. This trust is essentially an expectation – an assessment of what will happen in the future and in different contingencies. To the extent that individuals regard the rest of the world as unmotivated or unresponsive to their own choices, the economics of incomplete information are appropriate. To the extent that the results of individual conscious choices are seen as interdependent, the appropriate approach is game-theoretic.

As a first consideration, it is fruitful to distinguish trusting – whether to trust another entity (person, group, institution, etc.) – from trustworthiness (whether another entity should trust me). This distinction points the way for motivated and rational individuals to modify or facilitate the evolution of the rules of the game. In particular, choices of whom to interact (play) with and whose expectations to fulfil, disappoint or ignore determine the ‘network structure’ of the game.⁵

A second consideration is how the design of the game itself embodies trust. An important strand of the literature places trust in a contractarian setting. Bowles and Gintis (2000) relate trust to contractual incompleteness – which allows parties to economize on the information required to completely specify all contingencies and the obligations attached to them. As Bacharach et al. (2001) point out, this applies to default contracts – thus, trust is essential to the functioning of norms that allow markets to function⁶ Fukuyama (1995), Puttnam (2000) and Politt (2001) extend this, placing trust at the heart of social capital. This view is not uncontested.⁷ Some see an inherent conflict between contract (formal specification of transactional rights and obligations) and trust (informal or tacit surrender of powers of actions supported by expectations). Others make the pragmatic point that trust in incomplete contracts involves acting on incomplete information; some ‘trust-enhancing’ measures add information and thus weaken trust. Finally, the specific legal context of contracts provides for monitoring, verification and enforcement in the event of breach; this argues for looking at, for example, hierarchies of trust.⁸ These discussions go beyond the simple trust models considered here.

Trust is reflected in all the underlying data of the game: the set of ‘players’; their strategies or powers of action; their information; their motivations or pay-offs; and the solution concept used to summarize the information in the game.

Social games identify specific people ‘trusted’ to make decisions (play the game) and the decisions they are ‘trusted’ to make – drinking, driving, voting, making contracts, etc. This applies to acting on one’s own behalf, working with others, participation in collective activity and even (especially in the contract case) whether one is allowed to trust others or occupy a position of trust. Second, most decisions are made under uncertainty. In distinguishing objective from strategic uncertainty, and uncertainty from risk, we *trust* that objective uncertainty is not rationally-motivated and can be taken as exogenous, that other players are rationally motivated and can be understood by considering their preferences and beliefs and that important elements of these uncertainties can be reduced to risk⁹ and adequately described by equations and formal models.

Much recent work (including the present paper) represents trust as a strategic choice.¹⁰ An alternative aspect of trust is credibility – trusting information received and being believed by others (for example, threats and promises). Another strand of work (Bacharach and Gambetta 2001) attributes trust to motivational factors outside the rational behaviour framework. The foregoing considerations are summarized in Table 1.

Trust (trustworthiness and trusting behaviour) is valuable in complex interactive systems, but not necessarily good in aggregate. Rather, the *distribution* of trust can ensure sound expectations and an appropriate alignment of information, motivation and power to act. For this reason, a simplistic objective of maximizing trust can be myopic or even counterproductive. To clarify this observation,

it is useful to distinguish relations of trust between people, systems and organizations as shown in Table 2.

Table 1 Modelling aspects of trust

Aspect	Formal representation	Meaning
A decision to interact	Network structure	Choosing with whom to transact, trade or play.
Simplified societal mechanisms	Incomplete contracts, norms, conventions	Save 'costs' of covering all contingencies, potential 'partners'.
A collective good	Rules of the game, social capital, mechanism design.	The environment within which we negotiate and act
Being trusted	Players	Who gets to decide
Delegation	Strategies	What they can choose
Expected results of actions	Nash equilibrium	Trust in self-confirming beliefs or norms, even when held by others
Intention vs. accident	Uncertainty, rationality	Trust in models
A conscious decision	Trust as a specific strategy, precaution	Choosing to delegate, keep promises
Credibility	Revision of beliefs, incomplete contracts	Willingness to rely on information received
A private good	Reputation	Connection between what I say I will do and what I am expected to do, inference drawn from my actions.

Table 2 Trust and trusting

		Trusted party		
		People	Systems	Organizations
Trusting party	People	Societal trust (Fukuyama)	Agency, privacy, accuracy	Reputation, assurance
	Systems	Fault-tolerance	Complex system reliability	N/A
	Organizations	Agency	Reliance	Firm networking

Table 3 shows some advantages of coordinated trusting behaviour.

Table 3 Advantages of matching trust behaviour

		Trustworthy	Untrustworthy
Trusting	Appropriate delegation, specialization		Enforcement costs, costs of adverse incidents
Untrusting	Excess contracting, monitoring costs; race-to-the-bottom.		Lost gains from trade, inappropriate risk allocation

While it is advantageous, for example, for customers to trust e-commerce systems, it does not follow that more trust is better, or that it falls to firms or governments to build this trust, since due vigilance by customers is both empowering and efficient. On the other hand, many customers are unaware of the need for vigilance, or view the costs (including effort and knowledge acquisition) as too high. The analysis below (Section 2.2.5) recognizes that efficient vigilance may involve only a proportion of the population taking precautions and, hence, that the efficiency of the equilibrium outcome may be merely fortuitous – due to the natural tendency of individuals to avoid burdensome precautions and the (real or perceived) possibility of ‘free-riding’ on the diligence of others – in short, trust may be an under-supplied public good. Efficiency of risk allocation also

underpins policy favouring market competition over price regulation and negotiated reallocation of liability in tort systems.

On the narrow issue of trust between individuals and information and communication technology (ICT) systems, we may distinguish:

- Whether people trust ICT systems:
 - to act for them (agency);
 - with information about themselves (confidentiality, privacy),
 - to provide information that they can safely act on (accuracy, currency, authentication, identity, integrity, etc.).
- Whether trust among people or civil/private sector entities:
 - is helped or hindered by new systems;
 - improves or weakens approaches to crime (for example, neighbourhood watch).
- Whether trust between people and government reduces and improves the incidence of the impacts of crime and the burden of crime reduction.

This paper does not consider crime directly.¹¹ From the contractarian perspective, crime is breach of an (incomplete) social contract. The critical issue, addressed to some extent in the models in Section 2, is how individuals distinguish mistakes from crimes and whether they accurately attribute unanticipated and adverse events to specific actors and adjust levels of trust accordingly.

2 TRUST GAMES

2.1 Specific Game Theoretic Models

In this section, trust is analyzed through some specific game theoretic models. Table 4 summarizes these and the aspects of trust they are intended to illuminate.

Table 4 Trust games

Aspect of trust	Game(s)
Trust as a collective norm	Coordination ¹²
Fear of crime as distinct from lack of trust	Coordination and crime ¹³
Joining a network as a form of trust	Direct network games ¹⁴
Indirect vulnerability (the trusting nature of those we trust)	Indirect reliance
Need for vigilance	Interdependence
Social value of free-riding on costly vigilance	Vaccination
Optional availability of trusted channels or 'safe environments' and network externalities	Hybrid game
Trust is learnt from one's own experience and incomplete information about others	Sampling

Before proceeding, it is necessary to acknowledge specific features of some models: limited (usually pairwise) interaction; symmetry in payoffs and strategies; 'coordination' structure (players prefer to adopt the same behaviour *ceteris paribus*); the possibility of trust as equilibrium behaviour; 'rational actors' (self-regarding preference); and perfect information. At first sight, these seem highly restrictive, and the theorist's preference for the simplest possible model might seem inadequate. However, in addition to simplifying the exposition, these assumptions are not as strict as may at first appear; below, we discuss these assumptions and acknowledge alternative approaches. The Appendix briefly indicates the implications of the analysis for alternative formulations.

Trust in two-person settings differs in important respects from trust in multi-person settings. Trusted interactions expose players to indirect interactions with others and players engage in multiple pairwise interactions with a range of more-or-less identifiable partners – this is partially addressed in some of the models used in this paper (the indirect reliance, interdependence, vaccination and hybrid games). The third is beyond the scope of this paper - trust relations between groups and individuals or among groups may differ in kind from those among individuals –

for example, because group membership involves trust.¹⁵

Symmetry might seem to present a more serious limitation. In commercial transactions the roles of buyer and seller are not directly interchangeable (though typically conflated in general equilibrium models¹⁶). While buyers certainly have to consider whether to trust sellers, sellers may have far less need to trust buyers. This applies more to retail transactions and irrevocable sales of goods and services than to business-to-business (B2B) transactions and exchange of information in general communication networks. As ‘final-sale’ retail transactions increasingly give way to ‘lease’ arrangements (Rifkin 2000) liability does not pass irrevocably from seller to buyer. The continuing quality of the match between buyer’s needs and seller’s offer, resulting two-sided moral hazard and adverse selection problems and transactions and opportunity costs of dissolving relationships encourage the parties to consider trustworthiness before making a deal. This can already be seen in changing retail contracts; as public policy pushes back the boundaries of *caveat emptor*, limitations and protections built into commercial contracts (for example, software licences) increase despite empirical evidence that they are not very important to consumer choice. Section 3.4 applies the same methods to non-symmetric settings.

The coordination framework models situations where the parties must consider each other’s actions in choosing their own and where there are multiple equilibria. Anecdotally, it can be suggested that both high-trust and low-trust outcomes can be stable. This class of games includes those where equilibria can be Pareto-ranked¹⁷ and those where players prefer different equilibria.¹⁸ This paper primarily considers the Pareto-ranked setting and adopts the convention that the parties prefer the high trust to the low trust equilibrium. The analysis in Section 3.4 considers issues of horizontal trust in a collusive setting; this can give rise to coordination-game payoffs by including legal or reputation costs or retaliatory commitments. Specifically, a model of ‘price-matching’ oligopolists would have both a ‘competitive’ low-price equilibrium and a ‘tacitly collusive’ high price equilibrium preferred by all parties.

The key qualitative features of the coordination game (multiple, Pareto-ranked equilibria) apply even to the one-sided retail setting. There are always alternatives to e-commerce transactions, and often varying degrees of ‘depth’, which can make things two-sided. For instance, if an on-line buyer chooses to provide lots of information, the seller could realize further gains from trade via personalized transactions that benefit both of them. The seller does not ‘trust’ the buyer, but trusts the information provided enough to expend resources in devising and making targeted offers in the hope of future trades.

Of course, the stakes are not the same on both sides. Even when buyers and sellers reallocate risk and return through negotiation and market pressure, liability should depend on (asymmetric) ability to reduce and/or bear risk. But such payoff differences typically make only a quantitative, not a qualitative change to Nash equilibrium.

The games considered here model trust as a strict Nash equilibrium. An extensive literature¹⁹ on trust is built around an asymmetric game between a trustor and a trustee which does not have a trust equilibrium. In its starkest form, the trustor makes an unsecured transfer to the trustee – this amount is magnified (either in the transfer or by the trustee’s action) and a portion returned to the trustee. In the one-shot version of this game, played by purely self-interested rational actors, it is weakly dominant for the trustee to return nothing, so the equilibrium transfer is nothing as well. The inefficiency of this equilibrium can be remedied in various ways, for example, repetition, reputation, precommitment, third-party intermediation, auditing and other- or process-orientated preferences.²⁰ Most lead to equilibrium trust but retain the original, low-trust equilibrium. Because the former Pareto-ranked equilibria dominates the latter, they give rise to (possibly asymmetric) coordination-game ‘reduced forms’.²¹

The games also assume perfect information: the critical uncertainties concern the opponent’s action. This does not cover the trust issue as laid out in Section 1.2. Consider an ongoing interaction among players with different information. If player A does not do what player B expects B’s trust in the expected action was misplaced – B may respond by concluding that:

- A made a mistake (and B may form a belief about whether this will happen again).
- A was trying to send a message:
 - A may be sending a costly signal of willingness to switch to a better equilibrium;
 - A may be signalling private information ignored by the current equilibrium.
- B may have wholly misunderstood the game, or made some fundamental mistake in the

- chain of inference that led the current belief.²²
- A may be trying to convince B that B has misunderstood the game.

This raises intricate epistemological issues beyond our current scope, but a simple example (the ‘centipede game’) is briefly discussed in the Appendix. The basic contributions of game theory in this area are suitable ‘reduced forms’ for incomplete information games and more discriminating solution concepts that take account of knowledge, belief and inference.²³

The models developed here do not distinguish large and small players: all players are in the same situation as regards any other player to which a link is formed. This asymmetry is not addressed by rescaling, particularly not where link formation is concerned. However, the concept of trust as a convention (see Section 3) mitigates this in the asymmetric retail context, if the costs of contract negotiation favour use of standard-form contracts. Because such standard forms are incomplete on both sides and uncontracted contingencies are payoff-relevant for both parties trust is again two-sided; because they are standardized, there is at least a bit more symmetry across players and (for each player) across their partners.

The games considered here do not attempt formally to distinguish trusting from trustworthiness (Büschken 2000; Ben-Ner and Putterman 2001). Moreover, while the models are used to examine both trust as a norm and the indirect provision or erosion of trust through network interactions, they do not deal with third-party certification used to reinforce trust (or provide it more efficiently), but rather use a simple social capital metaphor. Other papers in this book consider social capital in more detail – for the present, it should merely be noted that it is particularly relevant in communication networks.

Finally, in recent years a richer theory of networked behaviour has begun to emerge which allows for greater asymmetry and richer structures of interaction (Page et al. 2002)). This points the way to a direct theory of trust relations involving more than pairwise interaction, and is clearly a fruitful area for further research.

2.2 Trust Games

In this section, we describe some simple games representing stylized views of trust and their associated equilibria. This style of analysis is suited to the static assessment of trust. We consider three classes of game: trust modelled directly as strategic behaviour; games that focus on network formation – and thus on linking as a form of institutionalized trust; and hybrid games combining elements of both approaches.

Direct coordination games. The simplest case is a ‘coordination game’ where players choose between high-trust and low-trust strategies. In addition to any interactions with ‘trusted’ systems and other parts of the environment where individual strategic considerations do not apply, they interact with their network neighbours. We assume that individuals choose a single strategy for all interactions; the more general case where strategy (trust) depends on reputation is considered below (and see Appendix 1 for mathematical details of models).

Strategies in this coordination game are complementary – player 1’s adoption of high trust increases player 2’s payoff from adopting high trust, and conversely for low-trust. The payoffs are thus assumed to be:

		Player 2	
		High trust	Low trust
Player 1	High trust	5, 5	A, B
	Low trust	B, A	3, 3

By complementarity $A < 3$ and $B < 5$ (there are positive network externalities; adoption of different trust levels is not profitable). Different interpretations of the strategies (and the parameters A and B) correspond to different notions of trust.

One formulation deals with accidents and uncertainty. In this view, the low-trust strategy involves costly precaution taken to stave off costly losses. Losses are also influenced by neighbours’

precautions (positively or negatively). In a second contractarian view, high-trust corresponds to a willingness to make (and entertain) many offers, to engage in incomplete contracts and to join contracts for complements. Offers from trustworthy people will be rewarding, relatively cheap to monitor and enforce, and mutually-reinforcing. The third interpretation focuses on trustworthiness: low-trust corresponds to sharp practice.

The parameter *A* is the consequence of trusting a low-trust person – if this means that advantages of trust are not realized or that one is taken advantage of, then *A* should be low. On the other hand, if a low-trust person takes extensive precautions that also provide benefits to his partners, then *A* should be relatively high. The parameter *B* is the consequence of not trusting a high-trust person. If this means foregoing the advantages of a mutually trusting relationship, or if the trusting partner also trusts untrustworthy third parties and thus jeopardizes the partnership, *B* could be low. On the other hand, if the partner's trust leads to more efficient delegation of work or provides indirect connection to mutually-trusting groups (whose interactions economize on monitoring and enforcement costs) then *B* could be relatively high. These interpretations are developed at greater length in Section 3.2 and especially Section 3.3, where different scenarios corresponding to these interpretations are considered.

All individuals prefer the high trust equilibrium, but they can get 'locked in' to the low trust outcome. If each player interacts with all others (fully connected network) the two possible (pure-strategy) equilibrium outcomes are homogeneous: all play H or all play L. If players are locked into the low-trust equilibrium, they can escape if a sufficient number jointly switch to the high-trust strategy – for instance, if they are temporarily indemnified against losses, or if policy measures reduce the cost of signalling (for instance, by raising *A*). In the formal equilibrium analysis, however, this requires a 'leap of faith' – even foresight will not necessarily induce signalling behaviour if the players take no account of the past. On the other hand, the less players discount the future, the more likely they are to try to escape the low-trust equilibrium by making unprofitable defections to the high-trust strategy. This is particularly likely:

- when one or more play as 'Stackelberg leaders' – assume other players will respond optimally to their choice of strategies; and
- when the geometry of connection creates 'hubs' – players to whom most players are connected by direct linkages or players whose behaviour is held up as exemplary. Hub players in such 'star-shaped' societal networks can be sure that changes will lead a critical mass of other players to re-examine their expectations.

The above analysis can be extended without much additional difficulty to distinguish fear of crime from lack of trust by adding a 'criminal' strategy (*C*). Assume that in the long run 'crime does not pay' – that is, H and L remain symmetric equilibria – but also that unilateral honesty does not pay when crime is conventional (*C* is also a symmetric equilibrium). The payoff matrix might look like:

	High Trust	Low Trust	Crime
High Trust	5, 5	<i>A</i> , <i>B</i>	-4, 3
Low Trust	<i>B</i> , <i>A</i>	3, 3	0, 0
Crime	3, -4	0, 0	1, 1

As before $5 > B$ and $3 > A$ and the same interpretations of *A* and *B* apply.

A game of indirect reliance.

The consequences of trust include the indirect impact of links to third parties trusted by the partner (see Appendix 1, Section A1.4). To examine this, we 'step back' from the specifics of network behaviour and simply assume that the ability to verify trustworthiness decreases as connections become less direct, since verification must rely increasingly on hearsay and reputation. In other words, the value of trust decreases with 'distance' (the number of intervening links). We further assume that direct links are costly to maintain.²⁴ It is natural to define a network as efficient if it maximizes total payoff to the players and Pareto efficient if any change makes at least one worse off. In this model, efficiency depends on the relation of link costs to value:

- if costs are high the only efficient network is the 'empty' one with no connections (or where

- no use is made of this form of interaction), corresponding to no trust;
- with intermediate costs star-shaped networks are efficient; and
- if costs are low the fully-connected network is efficient.

Players form trusted relationships where both players agree, and break them when at least one player wishes to do so. In the intermediate cost case, players only trust those that have trusted relationships with others, but no player wishes to trust more than two others in view of the costs. A network in which each player trusts exactly two others is a collection of separate rings. A ring cannot be part of equilibrium since each player could improve his payoff by breaking one link. The only equilibrium network (the 'no-trust' one) is not even Pareto optimal since a line offers both end and middle players strictly more.

A game of interdependence

A related game looks at the concentration of trust partners on each others' needs. For instance, the added risk or dilution due to 'outside entanglement' makes the value of trust between partners depend *inversely* on the number of trusted relations each party has and (because at least one should concentrate on protecting the relationship) inversely on the product of these numbers (see Appendix 1, Section A1.5). Again, equilibrium networks are typically inefficient. In simple cases, only the 'full-trust' network is efficient, but players could unilaterally improve by splitting into disjoint pairs.

A vaccination game

The previous games illustrate a basic 'paradox' of trust: each of us wishes to be trusted, but those who trust us might trust others whom we would not trust, or whose trustworthiness extends only to their direct partners (see Appendix 1, Section A1.6). This suggests that trust 'flows through' societal networks so there may be value in having some form of precaution to stop the spread of criminal or unreliable behaviour. This can be represented in a simple vaccination game: each player can choose whether to take a costly precaution. If either party to a pairwise interaction has taken precautions, both are protected; otherwise, both suffer. In a fully connected network the equilibrium level of p is only efficient if the level of cost is 'just right'. Network formation is rather trivial: any incautious player wishes to link to any cautious player and does not mind linking to other incautious players. Any cautious player wishes to link to all other players, so the fully-connected network is (at least weakly) an equilibrium. Clearly, this set up should be extended to a more realistic specification.

The hybrid game

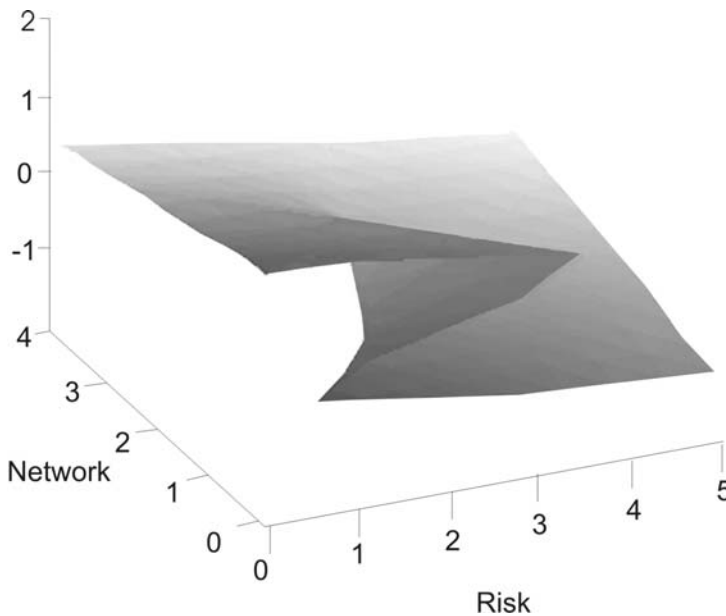
The 'value of trust' specifications (the indirect reliance and interdependence games above) can be interpreted as trust strategy games (the coordination games) by assuming that trust attitudes are prescribed or unobservable (and therefore equivalent to their expectations). All assume a single environment, where the payoffs can be interpreted in terms of the differential advantage to using the high-trust strategy (see Appendix 1, Section A1.7). This section uses this approach to consider a 'trusted' environment available alongside the 'game' environment. Players have unobserved differences in the degree to which they are willing to rely on the trusted channel. We begin by analyzing equilibrium behaviour in terms of underlying attributes of trust (risk and exposure). This strategic model is then compared with an incomplete information approach in which players learn by sampling the population; the two models can be distinguished empirically.

Description of the one-sided model. Consider a population of linked, but heterogeneous individuals, each of whom can choose between a high-trust and a low-trust strategy. Players evaluate the high-trust option according to 'net benefits' minus expected 'risk cost' (which includes any costs of switching to the high-trust strategy). The net benefits term is weighted by the player's relative preference or taste for the high-trust channel, which is not observed by others. It contains a fixed component (for example, reduced transaction costs associated with trusting certified public channels) and a term that varies with the number of other high-trust individuals with whom the player interacts and with the 'strength' of network externalities among players. The analysis assumes that idiosyncratic preferences are distributed according to a unimodal (single-peaked) distribution. This setup *can* give rise to multiple equilibria and this possibility *may* be related to underlying parameters representing the strength of network effects and the level of risk.

Equilibrium behaviour. Nash equilibrium (best replies to rational expectations about the behaviour of others) implies that the high-trust will be all those whose relative taste for the high-trust strategy is ‘sufficiently high’ – the cut-off value is increasing in risk cost and falling in net benefit. When network effects are small there is a unique equilibrium prevalence of trust (h), which is decreasing in risk cost and increasing in network effects. When network effects are strong there are three solutions (as shown in Appendix 1, under Equilibrium Behaviour). The highest and lowest solutions are stable and decreasing in both risk cost and network effects; the middle, unstable solution is increasing in both parameters. The overall picture is shown in Figure 1, with ‘trusting’ (h) measured on the vertical axis.

This picture illustrates the central characteristic of the model: when network effects are weak, trusting behaviour decreases continuously as perceived risk rises. As network externalities become more important, multiple solutions appear and gradually diverge. A similar result can be obtained from a model in which the risk parameter measures the likelihood of an abuse of trust and the network parameter is replaced by a measure of *exposure* or the likely loss from an abuse of trust.²⁵

Figure 1 Equilibrium trust as a function of risk and network externalities



A two-sided version. The above shows the dependence of equilibrium trust on perceived risk. To close the loop, we fix the strength of network effects and assume risk responds instantaneously to the prevalence of trusting behaviour. One extreme assumption is opportunism: risk rises or falls according to whether h is above or below a critical value h^* at which expected returns to abuse of trust just balance expected costs (including punishment). If network effects are small enough for trust to respond continuously to risk the model converges monotonically to h^* and the corresponding risk level. If network effects are ‘too large’ the system will cycle (clockwise) in a hysteresis loop.

The polar extreme is reassurance: risk falls or rises as h is above or below h^* , and the system tends to a high-trust or low-trust corner solution. The same result is obtained in Section 3.2 for evolutionary dynamics in a fully connected network (except that only one – risk-dominant – corner is stable). This result also emerges from a Bayesian model with partial adjustment of subjective risk estimates.

An incomplete-information model. Qualitatively similar results (roughly S-shaped time-paths for creation and erosion of trust) can be obtained from a simple incomplete information model where individuals sample trust through a random ‘word of mouth’ process. The dynamics depend critically on the credibility of this information (whether reports of the general level of trustworthiness are themselves trusted) and bias in the reporting of relevant information (for example, information selected by media or government to highlight negative or positive outcomes). In this model, high-

trust individuals at any given time are a representative sample of the full population rather than an ‘upper interval’.

Comparison. We can compare average trust per capita in the two models to track its response to a secular decrease in perceived risk brought about by, for example, government policy – which is equivalent to a fall in the critical taste parameter at which the individual is indifferent between high and low trust (see Table 5).

Table 5 Response of trust to fall in perceived risk

	Equilibrium	Incomplete information
No network externalities ($v = 0$)	Fall	Constant
Peer-to-peer ($v = 1$)	Rise, then fall	Rise

3 TRUST AS A CONVENTION

3.1 The Evolutionary Perspective

We now turn from the equilibrium to the evolutionary perspective. Evolution is the result of three coincident processes: variation, selection and heredity. As applied to conventions, variation results from natural mistakes: either conventional behaviour will be misinterpreted, or individuals will depart from convention. Selection is the result of mutual approval or rational recognition that the new behaviour may be better. Heredity is the result of societal learning and codification of behaviour into norms. To prevail, a pattern of behaviour should be self-reinforcing (or cohesive) and stable: variations should be small-scale, haphazard and not overly vulnerable to contagion.

Following the approach started by Kandori et al. (1993) and Young (1993, 1998), we consider a game played by a large, connected population, each of whom plays against all the others. For each potential convention (symmetric equilibrium), we use the size of spontaneous²⁶ deviations that could lead the players to wish to change their behaviour to assess its ‘resistance’ to contagion: the stable convention(s) are those against which other behaviours have the least resistance.

The following sections apply this to the direct coordination and criminal behaviour games described in Section 2.2 that are idealized representations of the ‘coordination’ aspect of trust. As argued above, the *qualitative* features of this representation are fairly general. The quantitative aspects (in particular the values of the parameters A and B) are not, so the following analysis considers how conventional behaviour varies with these parameters.

3.2 Stable Conventions in the Coordination Game

This section uses an evolutionary network approach to examine the emergence or disappearance of trust (see Appendix 1, Section A1.8). Each player plays against his or her ‘neighbours’ – the people to whom s/he is linked. Periodically, behaviour is reassessed and (subject to external shocks) revised. In consequence, actual behaviour depends on *cohesion* and *contagion*.

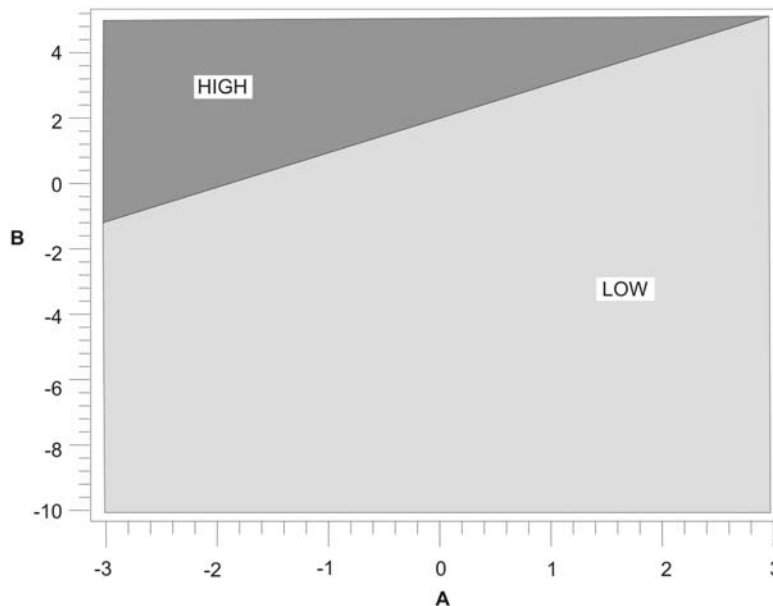
A fully-connected network

Here we use the direct coordination game from Section 2.2 – a player will want to switch to the high-trust strategy if a sufficient fraction ($\frac{3-A}{8-A-B}$) of his or her neighbours do. A fully-linked group

subject to random shocks will converge to high trust in the long run if this fraction is less than $\frac{1}{2}$ – in other words if high-trust is risk-dominant (better than low-trust against a random opponent). The low trust equilibrium can be stable even though it is not efficient. This simple model shows a ‘cheap’ way to escape low-trust equilibrium: in a non-evolutionary model, it would be necessary to indemnify victims of low trust by raising A to at least 3, whilst in the networked model it suffices to raise A to the lower level $B-2$. Figure 2 shows the stable conventions as a function of the

parameters A and B.

Figure 2 Stable conventions in the coordination game



The influence of geometry

The analysis extends to clusters or ‘small worlds’ where a limited set of close neighbours interact (Watts 1999; Jackson and Watts 2002). Given appropriate structure, such clusters can support stable diversity – both high- and low-trust conventions can prevail locally, even though only the ‘risk-dominant’ one would survive in a fully connected world. The same result is reached if the players are arranged in a circle; because each cares only about the actions of two neighbours, a player playing the risk dominant strategy would only switch if both neighbours switch to the other strategy, while one who joins his neighbours in playing the other strategy would switch if even one neighbour switches. The same is true in any regular network where all players have the same number of neighbours – a player is ‘less likely’ to switch away from the risk dominant strategy because this requires simultaneous deviation by more of his neighbours than a change towards the risk dominant strategy.

By contrast, asymmetric networks (such as star networks where all players share a common hub, as in the case of trusted services delivered from a central point) can support diverse outcomes. Consider three players linked to a common hub: each cares only about what the hub is doing, and the ‘spokes’ will copy any deviation of the hub. Any change by a spoke player in the direction of the risk dominant strategy will lead to all players eventually following suit. Of course, there are ‘more’ paths leading towards the risk dominant equilibrium, but where the risk dominant equilibrium is inefficient, this can be offset by farsightedness (Page et al. 2002) on the part of the hub player – or some suitable liability arrangement.

One final note concerns the speed of adjustment: as Ellison (1993) shows, local interactions accelerate the diffusion of conventions through a social network – small groups are relatively easy to saturate. Thus, ‘insularity’ generally means more rapid convergence to the risk-dominant equilibrium, and therefore may be regarded as good or bad depending on whether risk dominance coincides with efficiency.

3.3 Stable Conventions in the Face of Criminal Behaviour

In this game, the behaviours that might prevail in the long run in a highly-linked society – are the symmetric outcomes (see Appendix 1, Section A1.9). The stochastically stable conventions, following Young (1993), for games with more than two strategies are those for which the ‘resistance’ of each convention to shocks that might ‘tip’ it towards another convention is greatest.

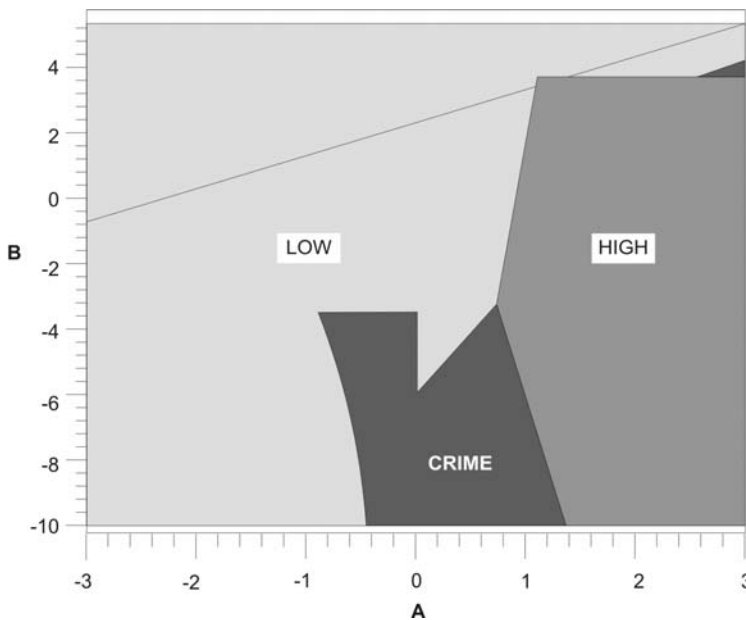
These shocks might be direct (for instance, if enough of one's neighbours switch from high- to low-trust, one will follow suit) or indirect (if an intermediate number of one's neighbours switches from crime to high-trust behaviour, one would wish to switch to low-trust behaviour, and the population as a whole might follow suit). To identify the stochastically stable convention, we sum the resistances associated with all transitions towards each convention; the stable convention has the lowest sum.

High trust is the unique stable convention if the payoff to low trust in partnership with a high-trust player (B) is small, but low trust may be stable if both B and the payoff to a high-trust player facing a low-trust player (A) are large. Roughly, B can be large if the precautions accompanying the low-trust strategy are relatively costless for players transacting with high-cost players, or equivalently where most of the benefits of a high-trust relationship accrue if at least one of the parties is trusting or trustworthy. A is large if the costs of signalling willingness to trust are relatively low.

It is again cheaper to stabilize high trust than to force unique high trust equilibrium. Network geometry matters – a densely connected network may require quite a few 'mutations' to flip behaviour, while a network of 'small world' neighbourhoods can 'escape' from the low-trust or criminal conventions relatively easily.

A parametric analysis of stability for different values of A and B is shown in Figure 3.

Figure 3 Stable conventions in the crime game



For the sake of comparison, the diagonal line shows the separation between low-trust and high-trust behaviour from the coordination game. We can make the following observations:

- In both games, low-trust behaviour is more likely as the consequences of trusting a low-trust person fall relative to the consequences of not trusting a high-trust person.
- Criminal behaviour can be stable under two circumstances: if the risks of being victimized by trusting a low-trust person are modest, but the loss involved in failing to trust a high-trust partner is substantial, or if there is no loss in trusting a low-trust person and a modest gain to a low-trust person when his or her partner switches from low- to high-trust.
- The introduction of criminal behaviour mostly works to destabilize high-trust conventions (except for the area above the diagonal line).
- Policies that try to stabilize high-trust behaviour by reducing B are less likely to work when there is a possibility of criminal behaviour – especially when A is low (that is, when there are substantial losses to trusting a low-trust partner).

3.4 Stable Conventions in Other Games

As mentioned in Section 2.1, the model used here is highly specialized. Alternative formulations can be used to investigate the robustness of the analysis and represent different aspects of trust. In this section, three alternatives are briefly contrasted to the symmetric direct coordination game introduced in Section 2.2.

	HT	LT		HT	LT
HT	(5, 5)	(A, B)	HT	(5, 10)	(a, b)
LT	(B, A)	(3, 3)	LT	(c, d)	(10, 5)
Symmetric coordination [$5 > B; 3 > A$]			Battle of the Sexes [$5 > c, d; 10 > a, b$]		
	HT	LT		HT	LT
HT	(e, f)	(5, 3)	HT	(e, f)	(3, 5)
LT	(3, 5)	(g, h)	LT	(5, 3)	(g, h)
Costly Precaution [$3 > e, f; 5 > h, g$]			Localized Precaution [$5 > e, f; 3 > h, g$]		

The Battle of the Sexes may be interpreted in terms of complimentary but costly precaution, differing levels of risk aversion or differing tastes for trust. Imbalanced levels of trust lead to, for example, litigation that is costly to both parties, but in the absence of litigation player 1 (2) prefers the high (low) trust outcome.

In the Costly and Localized Precaution games, exactly one player should take precautions; if both do the result is costly duplication; if neither does they suffer excessive damage. In Costly Precaution the careful player does worse (analogously to the vaccination game). In Localized Precaution vigilance is costless, but offers imperfect spillover protection to the other player (an extreme form of the indirect reliance game).

Exact numerical symmetry is irrelevant (for instance, the payoff to (HT, HT) in the coordination game could be changed from (5, 5) to (6, 8) without affecting the result. In each game payoffs can be modified so that the equilibrium payoff set is asymmetric).

Section 2.2 showed how players could get locked in to the inefficient low-trust equilibrium. In the Battle of the Sexes, pure-strategy equilibrium necessarily favours one or the other player, even if, for example, $a=9=b$, in which case (HT, LT) would be both equitable and Pareto optimal – but not an equilibrium. The Precaution games have asymmetric pure strategy equilibria. The disequilibrium low trust outcome may be Pareto optimal in Costly Precaution and the disequilibrium high trust outcome may be Pareto optimal in Localized Precaution.

The more interesting results concern the stable conventions. The following summarizes the conditions under which high trust is stable.

Game	High-trust player(s)			
	<i>Both</i>	<i>Player 1</i>	<i>Player 2</i>	<i>Neither</i>
<i>Symmetric coordination</i>	$2 > B-A$	NA	NA	$2 < B-A$
<i>Battle of the Sexes</i>	$a-c > 5 > b-d$	NA	NA	$b-d > 5 > a-c$
<i>Costly Precaution</i>	NA	$h-f > 2 > g-e$	$g-e > 2 > h-f$	NA
<i>Localized precaution</i>	NA	$h-f > -2 > g-e$	$g-e > -2 > h-f$	NA

4 CYBER TRUST AND CRIME PREVENTION AND MARKET STRUCTURE

4.1 Structure, Conduct and Performance

Trust is essential to commercial transactions where costs of contractual completeness are high or legal frameworks for contract enforcement are unreliable. Trust among firms has traditionally been associated with collusion, but increasingly (in the so-called new-economy) also with informal 'networking' arrangements that economize on transactions and communications costs and improve collective efficiency. Trust between consumers and firms also provides transaction cost savings and *may* provide incentives for competition and innovation leading to improved price and quality, but equally, trust constitutes a barrier to customer switching (to the extent that it is relation-specific) and thus weakens competition.

This section addresses two connections between industrial structure, conduct and performance and the constellation of issues around cyber trust and crime prevention; the impact of market failure on levels of trust and the reciprocal impact of trust on market structure and conduct.

4.2 Competing for Trust

Structure, conduct and performance (SCP) analysis can usefully be applied to these issues in several ways. The literature has extensively developed the theory of reputations in terms of both price and quality. A firm's market power may derive from its reputation; conversely, market power can allow the firm to signal its quality and build reputation more effectively. Because reputation and trust influence consumer decisions, they in turn alter incentives to use trust as a way of creating or consolidating market power.

Reputations

The most obvious representation of trust in the analysis of commercial transactions is the game-theoretic treatment of reputations. The starting point is the existence of relationships where the parties' interests may diverge, and where informational asymmetries and/or prohibitive costs of reliable contractual protection make it difficult to reap the benefits of interaction without informal trust-like mechanisms.

Horizontal trust

Such mechanisms are used among groups of horizontally related firms that trust each other to cope with an uncertain world or to counter specific challenges. The object of such trust may be market allocation, prices and quality of even the provision of information. Firms wishing to reach agreement on these strategies in pursuit of mutual interest face legal obstacles to formal contracting – hence the uses of the term 'trust' to refer to cartels at the turn of the 19th century. To overcome this inability to contract and the combination of temptation (profitable unilateral defection) and exposure (negative externalities), firms have developed a range of trust-enhancing strategies, including social contacts, interlocking directorates, most-favoured customer clauses, etc. As a matter of definition, we should regard practices that facilitate collusion as 'trust based' to the extent that they are not directly enforceable, or sanction activities that cannot be directly verified. The cyber environment modifies the received thinking in a variety of ways. Most obviously, firms competing in electronic marketplaces have expanded opportunities for using anonymity to cloak departures from collusive agreements and, in some cases, a global platform for their activities. This could be seen as increasing the likelihood of defection – and thus for an increased need to rely on trust. The same factors also suggest an expanded scope for market-sharing agreements²⁷ as opposed to classical price fixing. A second cyber influence is the scope for rapid and effective response to defection and the relative ease with which extensive information could be provided – this may enhance the extent to which firms are able to trust each other. Finally, the emergence of new forms of market contact²⁸ that reduces search and transactions costs, which may simultaneously affect the potential impact of defections on other cartel members, influences the likelihood of effective detection and the power, speed and credibility of retaliation.

Vertical trust

Trust may also be critical in vertical relationships, such as input markets, providing access and retail or sales to consumers. Here, the impact of cyber technologies on trust and market power can

be seen at various stages of the transaction: search, payment, fulfilment and follow-up.

At the *search* stage, new technologies can provide more information and 'broader' information (for example, quality attributes or other non-price comparators). This, in turn, can sharpen consumer search – at least up to the point where effective comparison becomes impossible due to a glut of rapidly changing information. This could lower the cost of differentiating products from each other (rendering competition monopolistic and inefficient); alternatively, it could create a niche for information intermediaries (such as comparative search engines) that sharpen price competition – typically at a modest cost in deadweight loss. However, it is not obvious that the 'power' of search engines is entirely benign. While they do aggregate comparative information, they also order it, and there is no *a priori* reason to suppose that the ordering serves the interests of some (or even all) consumers, or provides open access to consumers by small-scale producers. In terms of trust, the question raised is the following: if consumers are inclined to trust large firms on the assumption that untrustworthy firms (or products) could not sustain prominence in the market, under what conditions should they trust the information intermediaries responsible for that prominence – since typically search engines assume no liability for the accuracy of the information provided to them or for the quality of goods purchased on the basis of this information (even in the case of electronic markets such as eBay) when intermediation extends beyond information to brokerage. The information itself may involve (or give the impression of involving) less commitment.²⁹ In particular, whether customers will trust comparative search engines to identify good buys³⁰ and whether the excess of information thus provided increases or reduces customers' feelings that they have selected a best buy³¹ are matters that must be determined empirically and, as yet, the theory of vertical separation between sellers and information intermediaries is in its infancy (Bailey 1998; Bakos 1997). Certainly, there is clear evidence that markets for information are even more prone to concentration than the up- and down-stream markets for 'real' goods and service (Brynjolfsson and Smith 2000). At this stage, it seems reasonable that the new technologies are more likely to enhance trust for standard commodities than for highly personalized or differentiated ones.

In cyber environments, *payment* typically involves one or more financial intermediaries. The most critical aspects of trust here are connected with the provision of (limited) access to financial assets and information; both parties need to be assured of each other's identities, to have confidence that payment and fulfilment will proceed in real time once the electronic transaction is completed, and that opportunities for repudiation are limited to those permitted in the contract. In addition, the paying party will need assurance that financial information will not be used to make further, unauthorized transactions. These issues have been discussed at length in the literature. We merely point out that, in contrast to product search – where multiplicity and competition are the most trusted sources of assurance – in the payment area there are advantages to prominence and a certain degree of concentration. However, with prominence may come increased risk – consumer concerns include dangers from mistakes, malfeasance by trusted parties or intermediaries and insufficient precaution in guarding transaction details – and insufficient precaution, at least, may increase as a financial service provider's reputation and market share increase, if these signs of success make it a more attractive target.³²

In the *fulfilment* phase, the impact of the new technologies is felt primarily through the globalization of commerce – the selling party may be located in another (even an unknown) jurisdiction, and pursuing consumer rights may be difficult or expensive. It may be that the prominence that comes with a dominant position can tie performance to reputation sufficiently to reinforce trust.

As regards *follow-up*, the analysis of reputations and quality emphasizes the importance of *signalling*. In particular, high quality can be signalled (and specific trust enhanced) by providing verified information (for example, quality certification by independent third parties) or assurance. The efficacy of the former strategy may be limited by psychological factors – in particular, provision of information relating to risk may heighten risk awareness. This is particularly true where the effect of providing information is to shift either the effective or the perceived allocation of risk – typically, decisions and consent have more weight and attract more liability the greater is the information provided in advance.

In oligopolistically competitive environments, firms may attempt to signal their relative trustworthiness by talking up problems encountered by their competitors, but this may reduce trust in the market as a whole, particularly when the problems arise outside or among the firms. Certification is an attractive alternative, but depends again on the reliability of the certifying authority. Much of the recent literature on, for example, cyber-notaries or Internet governance is

concerned with the relative merits of competitive and coordinated certification, and it is fair to say that at the moment the question remains unresolved.

One final comment concerns the evolving nature of goods and services delivered over the Internet – increasingly, these have a heavy information content and run into the classic problem of asymmetric information – the buyer and seller cannot assess the suitability of the match between the supplier's offering and the customer's need without an exchange of information, following which the buyer (or indeed the seller, in the case of personally identifiable information) has an incentive to exploit the information without commensurate payment. To fit such transactions into the relatively anonymous framework of retail commerce, a fair amount of trust is required – perhaps on both sides. If the information is to be re-used in some way, a further question arises regarding the allocation of any resulting loss (or gain). Put simply, it may be difficult to set appropriate limits on the provenance or use of information without introducing such heavy informational and contracting requirements that mutually beneficial transaction may not occur.

This is analogous to the agency situation pertaining between patient and health care provider – without an exchange of information (the patient describing symptoms and the provider describing treatments) effective treatment decisions cannot be made by either side. The traditional approach to preventing abuse in this situation of local monopoly³³ is to use a one-sided liability approach – trust is tied to certification of health care providers, backed up by the apparatus of tort law. However, these arrangements have not survived technological advances, including the availability of more effective treatments (which providers may not have incentives to provide) and increased consumer information (in no small part due to health-related information available on the Internet). The point of direct relevance is that the provision of information to the 'principal' (the patient) has undermined the former relationship of trust; patients no longer trust providers' expertise and providers no longer trust patients to follow advice or refrain from litigation.

Network externalities

The third type of relationship connects economic entities that produce complements rather than substitutes or inputs. The world defined by new technologies is increasingly a networked one. The simplest view of networks is of groups – trading partners, users of specific products, members of communities. Complementarity is perhaps the defining characteristic of such group-orientated economic relationships (Katz and Shapiro 1994). A more sophisticated view considers the geometry of networks, distinguishing indirect from direct connections (and thus potential from immediate interactions). In this approach, the very act of joining a network involves trust – the joining party must trust that indirect connections will not damage his interests, and incumbents must trust each other not to form damaging links. To make matters concrete and focus on market failure, let us consider the provision of software services, which may include some degree of security. For informational goods marginal costs of production are very low and the initial purchase or decision to adopt may involve a commitment to further purchases or adoption of software products – often from other producers. The net present value to the firm of its customer base equals the customers' aggregate switching cost (Anderson 2003). The impact of competition is that incumbents try to maximize and potential entrants try to minimize switching costs. Both types of strategy impose costs on the market as a whole. Moreover, 'churn' can undermine trust in the stability of the market and reduce suppliers' incentives to invest in durability and continuity. As applied to complements, this further implies that incumbent firms have incentives to 'lock-in' suppliers of complementary products as a way of ultimately locking in consumers. In the specific context of security, this may mean lowering interoperability hurdles for 'compatible' products – effectively using the potential for attack at the interstices between software products as justification for extending the 'trusted zone' to enclose producers of complements.

Another implication of the network perspective is that these software products are used to mediate transactions and communication between other people, so the creation of proprietary standards builds a network externality among users of extended systems – the value to each person of using the system increases with the proportion of his or her contacts who use that system. Again, the strategy of a dominant incumbent is to attempt to maximize such network externalities. This in turn leads to U-shaped adoption curves, 'tipping equilibrium' (capture of the whole market by a single 'extended standard') and to abrupt jumps and local irreversibility leading either to cyclic variation or to dominance that endures radical changes in technology.

The implications for market structure are that the first-mover advantage and the need to capture

suppliers of complements lead firms to reduce security barriers to developers, to share information with them and to shift the cost, complexity and liability burdens of security to customers. The implications for trust depend on the ability of customers to determine whether security provisions are effective (and appropriate). If customers cannot identify 'good' security, a form of Gresham's Law will operate, and inferior security precautions will drive out good ones. Continuing demand must then be sustained by exaggerating threats. Finally, it should be noted that even effective security precautions may merely displace risk.³⁴

Consumer protection

Quality can also be signalled by assurance – usually taking the form of a warranty or compensation scheme tied to breakdowns of trust. In contractarian terms, this is analogous to penalties or liquidated damages.

At the most practical level, work on product safety and reliability has distinguished two strategies that can be used to signal quality and build trust – despite being opposites, both may coexist in equilibrium. A firm wishing to convince customers of the quality of its products can provide either extensive or minimal warranty protection. In the former case, the firm credibly signals product quality because the warranty would be too expensive to offer if quality were poor. In this case, the customer does not need to trust the firm,³⁵ but the firm may need to trust the customer not to make frivolous claims. A well-known firm can also credibly signal quality by providing unusually low levels of protection, since it is placing its reputation on the line by doing so.³⁶ In this context, we are essentially dealing with trust in the informational sense, since the consumer does not have much control over the extent of risk transfer, merely the choice of whether to buy (or lease) or not. Note also, that the signalling value of trust-enhancing actions is relative to expectations – in other words, to the prevailing conventions regarding behaviour. This will be taken up in the next section.

Liability

A final area where the industrial organization perspective is relevant is the economics of liability rules. Wherever trust and crime are important there are, by definition, externalities, which can be mitigated by precautions taken by the affected parties. The recommended collective response (in civil, criminal or contract law) is to allocate liability for consequences. If negotiation is costly, the allocation must balance welfare considerations of efficiency and equity (here, fairness). In some cases, the externality seems to be fairly one-sided (in other words, it is for me to trust you rather than for us to trust each other) as are the prevailing liability rules (*caveat emptor* or the complementary strict product liability). According to standard practice³⁷ for efficiency's sake (and 'natural justice') liability should fall on the least-cost avoider (sometimes least cost insurer) of a particular harm. In cases where both parties make substantial contributions, or where avoidance costs are fairly symmetric, the same argument gives rise to the sort of 'relative negligence' rules used in transportation settings (for example, automobile or marine liability). An extreme case is the 'eye-for-an-eye' rule in which all parties bear full liability: this ensures efficiency *ex ante* (people take efficient precautions), but is not efficient *ex post* (people collectively pay the damage several times over). This case corresponds to a situation of criminal law – the criminal is penalized for his actions, but the victim is not (or not fully) indemnified for her loss. If parties can bargain costlessly over liability and if the total loss taking into account the liability system does not exceed the original harm then efficiency can be achieved regardless of how liability is assigned – but in this case rules like the eye-for-an-eye rule lead to too much precaution.

These considerations seem particularly applicable to risks such as those associated with computer viruses, spam and harmful or illegal content, since a wide range of people can take precautions with varying degrees of efficiency. In addition, precautionary activities themselves have externalities: some have the effect of protecting others (for example, by shutting down the offending communication at the source), some do not affect risk to others (for example, protecting one's own machine) and others may even transfer the risk or costs to others (for example, refusing to communicate with those who appear to have sent offending messages, but whose identity has been stolen and then used to send such messages).

These considerations loop back into industrial organization because the 'ownership' of risks and precautionary activities – and thus the trust placed in the system by participants (whether end consumers or B2B partners) – are affected by the degree of monopolization, the prevalence, adequacy and ownership of standards, and the 'networking' among market participants. This is a

two-way connection: industrial organization economics can be directly applied to the study of trust and crime – especially in commercial relationships – and considerations of cyber crime and trust can focus the development of industrial organization to deal with a range of ‘new economy’ issues. One specific area of application might be the economics of information assurance.

5 CONCLUSION

5.1 What Does Economic Theory Tell Us About Trust?

Economic analysis tends to endogenize trust – to treat it as an aspect of the functioning of economic systems whose roots are hard-headed calculations far from the intuitive notion of trusting behaviour. The individual’s decision to trust or be trustworthy is analyzed in terms of expected costs and benefits. Thus, person A will trust person B to capture gains from trade; B may have more relevant information, or greater powers of action, or the costs of fully negotiating and specifying the joint activity desired by A and B may be too great. A must assess both his exposure and B’s temptation, and decide *not* to use all relevant information or to proceed in the absence of relevant information.³⁸ From this perspective, trust and identity services (as provided by, for example, cyber-notaries and trusted third parties) are valuable products of the economic system. Because such entities provide information goods the relevant economics is the analysis of incomplete information. These models show that institutions that permit credible or verifiable signals (assurance) and informal institutions (for example, reputations) can improve efficiency; and that specific contractual forms can align incentives in cases of hidden information (adverse selection) and hidden action (moral hazard).

But trust is also a public good. Partially, this is because the acts of trusting, refusing to trust, fulfilling promises and breaking promises have both positive and negative externalities, from which it may not be possible to exclude others. In addition, it may not always be possible to distinguish individuals, or to fully anticipate the vulnerabilities to which a decision to trust exposes one – this is particularly true of the complex transactions of the cyber world. Thus, a person cannot fully ‘own’ trust or exact payment for it and it is possible to ‘free-ride’ on the trust or trustworthiness of others. To the extent that trust is costly, it will thus be underprovided.

5.2 Comments on the Analysis

The game theoretic analyses shed light on trust as a public good (especially the vaccination game). Efficient allocation depends on both the total and the incidence of costs and on the power to avert damage (or contribute to trust). Some of these can be measured, aggregated and traded, but others cannot. Other problematic possibilities suggested by the vaccination analogy include free riding (the ‘tragedy of the commons’ if trust is congestible) and enclosure (the creation of trust ‘clubs’ from which others are excluded).

The analysis has also highlighted some common features that may be more broadly applicable. Chief among these is the S-shaped adoption curve, which also occurs in another guise in the discussion in Section 4.2 with respect to competing for trust. The curve reflected the impact of interoperability, and described the time-path of adoption of a trust-enhancing technology: initially, when adoption is low, the interoperability advantages are also low, and only early adopters use the new technology. As the number slowly grows, so do the advantages of joining, and the rate increases, until virtually the entire ‘group’ has joined. This sort of behaviour results from two dynamic forces: the increasing strength of network externalities and the decreasing pool of ‘outsiders’ available to join. Superficially, the adoption of societal norms, the formation of expectations (for example, of trustworthiness or reciprocity), the accumulation of market power and the accumulation of experience (and the related concentration of a market ruled by an ‘experience curve’), would seem to follow a similar pattern. However, the S-shaped pattern runs deeper still – the time-path is based on the assumption of convergence to a prevailing norm. In the cyber trust world, this may apply more directly to the small-scale dynamics of the formation of specific groups and oscillation (as illustrated by the hybrid model) between high and low levels of trust. Throughout, the perspective has been on the evolution and dynamics of trust rather than on its static or equilibrium properties.

Another general conclusion is that there may well be a tension between stability (equilibrium) and efficiency. This may seem surprising from the standpoint of competitive economics, but is

consistent with both the view of trust as a matter of externalities and with a series of results showing the generic inefficiency of (usually mixed-strategy) Nash equilibrium.³⁹

From the SCP point of view, the need to rely on trust can reinforce the ‘tipping’ tendency towards market dominance and thus limit the effectiveness of competition. Similarly, competition to provide trust (or trust-enhancement and assurance) can compensate for public good under provision of precaution, the natural disinclination of many stakeholders to take appropriate precautions, and the consequent inefficient allocation of risk. In information markets, the struggle to lock in customers can affect quality, the extent of information assurance and information security – for example, when bottleneck suppliers of ICT services or software reduce internal security barriers to ‘capture’ producers of complementary products and impose those costs instead on consumers.

Because trust is bound up with expectations, incompleteness of information, of markets and of contracts is critical. This suggests a role for policy that supports self-regulatory mechanisms (for example, open standards, reputations) or provides for appropriate allocation or low-cost trading of liabilities. Other aspects of trust enhancement should be provided by public bodies or not-for-profit, open self-regulatory bodies. Because technology strongly affects trust, but is also liable to foreclosure through intellectual property rights systems, the Terms of Reference should probably favour institutional over technological specificity. Ownership of technical standards, while both inevitable and appropriate, should carry a ‘price’ in terms of money, liability or responsibility – the role of government may simply be to oversee the markets on which this price is determined and paid and/or to certify that the price has been paid. Finally, the sustainability of trust relationships may depend on asymmetry among the participants – in such cases, ‘improvements’ that reduce this asymmetry (for example, the provision of identical information to both sides) may actually undermine trust.

APPENDIX 1 MATHEMATICAL DETAILS OF MODELS

To simplify the exposition, mathematical details of some of the models are included in this appendix.

A1.1 The Fundamental Trust Game and its Reduced Form

A ‘fundamental’ trust game used in experimental and theoretical analyses (see, for example, references in note 16) involves a trustor with wealth of, say £10, who allocates to a trustee $\sigma_1 \leq 10$. The money is doubled en route by the effort of the trustee, who can then decide to return $\sigma_2 \leq 2\sigma_1$ to the trustor. The payoffs are $(U_1, U_2) = (10 - \sigma_1 + \sigma_2, 2\sigma_1 - \sigma_2)$ – the only subgame perfect equilibrium is the inefficient no-trust outcome $(\sigma_1, \sigma_2) = (0, 0)$. However, there are many ways out of this ‘trap’ – for instance, if player 2 (the trustee) can precommit to a strategy, efficient outcomes become far more likely – specifically, suppose $\sigma_2(\sigma_1) = 2\alpha\sigma_1$ if $\sigma_1 > s_1$ (and 0 otherwise). The trustor’s best reply is shown in the following matrix.

	$\alpha < 1/2$	$\alpha = 1/2$	$\alpha > 1/2$	$\alpha = 1$
$s_1 = 0$	No trust	No trust or any amount $\geq s_1$	Any amount $\geq s_1$	Any amount $\geq s_1$
$0 < s_1 < 10$	No trust	No trust or any amount $\geq s_1$	Maximum trust	Maximum trust
$s_1 = 10$	No trust or minimal trust	No trust or any amount $\geq s_1$	Maximum trust	Maximum trust

From this, it is clear that if the trustee precommits to $\alpha = 1/2$ the trustor is indifferent as to how much he commits, so the trustee should further stipulate $s_1 = 10$.

If we fix the player’s strategies in advance (so that (σ_1, σ_2) are constants), and impose feasibility we obtain the following 2x2 game:

		Trustee	
		σ_2 (fulfil)	0 (violate)
Trustor	σ_1 (trust)	$10 - \sigma_1 + \sigma_2, 2\sigma_1 - \sigma_2$	$10 - \sigma_1, 2\sigma_1$
	0 (withhold)	10, 0	10, 0

If $\sigma_1 < \sigma_2$ we get the 'standard trust game' in which fulfilled trust is not an equilibrium (because the violate strategy is weakly dominant). If $\sigma_1 > \sigma_2$ we get a weak prisoners' dilemma. A staggered prisoners' dilemma can give rise to the standard trust game if appropriate tit-for-tat (TFT) strategies are used. Indeed, we can interpret trust as cooperation and apply Axelrod's (1984) experimental results. For instance, if players remember only the other player's previous move, successive elimination of weakly dominated strategies eliminates all but two strategies for each player; tit-for-tat (trust/fulfil at the beginning, then trust/fulfil if the opponent played fulfil/trust last time) and the 'greed' (G) strategy (always withhold/violate). Using the long-term average payoffs corresponding to full cooperation and equal division of gains ($\sigma_1 = 10, \sigma_2 = 15$) in the first matrix below, tit-for-tat weakly dominates greed. However, this rests on the assumption that players facing greed are indifferent between their strategies. An alternative would be to modify the payoffs slightly to represent psychological attitudes. The middle game below embodies a feeling of righteousness that comes from behaving in a reasonable way against an intransigent opponent. This is not a naïve reliance on trust, since the tit-for-tat strategy does meet like with like. The first game represents an attitude of 'fool me once, shame on you; fool me twice, shame on me'.

		TFT	G
TFT	(15, 5)	(10, 0)	
G	(10, 0)	(10, 0)	

Weak dominance

		TFT	G
TFT	(15, 5)	(10 + ϵ , 0)	
G	(10, ϵ)	(10, 0)	

Relative righteousness

		TFT	G
TFT	(15, 5)	(10, 0)	
G	(10, 0)	(10 + ϵ, ϵ)	

'Fool me once'

Relative righteousness makes TFT dominant – at least in a world with only TFT and G. By contrast, the 'fool me once' game is a coordination game – and TFT-type trusting behaviour is risk-dominant (stable in a symmetric network) if $5 > \epsilon$.

A2 A Simple Trust Dilemma in an Extensive Form Game

One simple illustration of the complexity of trust is the question of rational play in the centipede game:

- Player A is offered repeated (say 100) opportunities to end a game.
- If he ends the game at the first opportunity, players A and B each get £1,000,000
- Each time A lets the game continue, this amount decreases by £10,000; thus one the 100th opportunity ending the game gives each player £10,000
- If A allows the game to continue on the 100th round, the choice of whether to continue passes

to player B. If she ends the game, each gets £200; if she allows the game to continue, player A receives a final choice: either each gets £500 or each gets £10.

The problem for player B is that, if she *trusts* in 1's rationality, she should allow the game to continue, thus getting the same £500 A will choose 'for himself.' On the other hand, the mere fact that she has to move is evidence that player A has deviated from rationality not once, but 100 times. Should she 'trust' the internal logic of backward induction or the external logic of the game? It could be, for instance, that A puts lexicographic priority on prolonging the game (which he enjoys) and only then wants to maximize payoff – in which case she should allow the game to continue (while wishing, perhaps, that he was more mercenary). On the other hand, it could be that A does not care about money at all – or actively hates it, or wants to minimize B's payoff, all of which should lead her not to 'trust' A with a further move.

A3 Costly Alternatives to Trust

One final example of trust in an imperfect information setting examines a situation in which the trustor does not necessarily know whether his trust has been betrayed – in this setting we can ask 'how far' the trust extends; in other words the form of the contract between trustor and trustee.⁴⁰ The example is a modification of the fundamental trust game of Section A1.1. The trustor can give a fixed amount to the trustee, which increases to a random amount $I(\omega)$, where the 'state' ω is observed by the trustee but not by the trustor. The trustor believes that the state ω is distributed according to a (common-knowledge) prior distribution $p(\omega)$. To keep things simple, assume that the states are ordered in such a way that I is increasing in ω . The trustee 'should' return an amount $R(\omega)$ to the trustor. Even if $I(\omega)$ is always bigger than the original transfer ω , this game shows the basic trust problem: the trustee will always return the smallest possible amount – the amount corresponding to the worst outcome. But suppose that the trustor can 'audit' the state for a payment of $\alpha > 0$ and thus enforce the contract. To complete the model, let us assume that both parties are risk-neutral, and that the trustee cannot be forced to pay more than $I(\omega)$. The 'trust contract' specifies the set of states A in which the trustor calls in the auditors and the amount $R(\omega)$ that the trustee is supposed to return to the trustor. Finally, suppose the trustor can design an optimal contract – one that maximizes the trustor's payoff subject to the participation of the trustee and the limitation that auditing is costly. Alternatively evolutionary pressures (competing trustees) may lead to an optimal contract.

Suppose the true state is ω ; the trustee will only report a different state ω' if this does not lead to an audit (ω' is not in A) and if it does increase the trustee's payoff ($R(\omega) > R(\omega')$). Therefore, any two states not in A must stipulate the same repayment. Denote this constant repayment as R^* . It is easy to show that in the optimal contract, A is precisely the set of states where the trustee cannot afford this repayment (that is, where $I(\omega) < R^*$), and moreover that when the auditors are called in, the trustee should be made to return the entire remaining amount ($R(\omega) = I(\omega)$ for ω in A). This means that the stipulated repayment, R^* completely determines the trust contract. A higher R^* leads to more frequent auditing; a lower R^* leads to a smaller return on trust. Ignoring the trustee's willingness to participate, there is thus an 'interior' level of R^* . If the trustee has a lucrative 'outside option' the actual level of R^* will be determined by the condition that accepting the trust compensates the trustee for foregoing the outside option. This model can be expanded to take account of differences in the productivity of different trustees, the interaction among trustors (in other words, a given trustee's outside option for accepting one person's trust is, at least in part, determined by the reward for accepting another's), 'distance-related' costs (hence networking) and changes to the auditing cost (reflecting, for instance, different costs of monitoring and enforcement in an electronic world).

A4 Game of Indirect Reliance

Let N be the set of players, $N_i(\Gamma)$ those directly trusted by player i in network Γ and $d(i,j)$ the minimum length of a path from i to j .⁴¹ The benefit to i of connection to j is:

$$b_i^j = \begin{cases} \delta_{ij}^{d(i,j)-1} - c_{ij} & \text{if } j \text{ is connected to } i \\ 0 & \text{otherwise} \end{cases}$$

The basic value, δ , of trust combines both the value of the exchange and the assurance that the other party is trustworthy. We assume $\delta < 1$, so the expected value of indirect trust falls with distance. The value to i of a network Γ is thus:

$$v_i(\Gamma) = \sum_{j \in N} \delta_{ij}^{d(i,j)-1} + \sum_{j \in N_i(\Gamma)} c_{ij}$$

A network is *efficient* if it maximizes the total payoff of the players and *Pareto efficient* if any change makes at least one player worse off. Efficiency depends on the relation of links' costs to value:

- If costs are high ($c > \delta + \delta^2/2$ in a setting with 4 individuals), the only efficient network is the no trust 'empty' network.
- With intermediate costs ($\delta < c < \delta + \delta^2/2$), star-shaped networks are efficient.
- If costs are low ($\delta > c$), the fully connected network is efficient.

Players form trusted relationships where both players agree, and break them where at least one player wishes to do so. In the intermediate cost case, players only trust players who have trusted relationships with others ($\delta < c$), but no player wishes to trust more than two others in view of the costs. A network in which each player trusts exactly two others is a collection of disjoint rings – but a ring cannot be an equilibrium since each player gets $2\delta + \delta^2 - 2c$, but could improve his payoff to $\delta + \delta^2 + \delta^3 - c$ by breaking one link. The only equilibrium network is the 'no-trust' one, but this is not Pareto optimal let alone optimal, since a line offers each end player $\delta + \delta^2 + \delta^3 - c > 0$ and each middle player $2\delta + \delta^2 - 2c > 0$.

A5 A Game of Interdependence

The value of trust between i and j depends *inversely* on the number of trusted relations each player has (for instance, because 'outside entanglement' brings added risk) and (to reflect the need to have at least one party concentrating on protecting the relationship) inversely on the product of these numbers:

$$b_i^j = \frac{1}{\#N_i(\Gamma)} + \frac{1}{\#N_j(\Gamma)} + \frac{1}{\#N_i(\Gamma)\#N_j(\Gamma)},$$

so

$$V_i(\Gamma) = 1 + \sum_{j \in N_i(\Gamma)} \frac{1}{\#N_j(\Gamma)} + \frac{1}{\#N_i(\Gamma)\#N_j(\Gamma)}$$

Again, equilibrium networks are typically inefficient: the only efficient network in a four-person game is the 'full-trust' one (in which $\#N_i(\Gamma) = 3$ for all players), which pays each player $7/3$. If the players broke into disjoint pairs (so $\#N_i(\Gamma) = 1$ for all players) they would each get 3 – truly a case where 'two's company and three's a crowd'.

A6 A Vaccination Game

Each player chooses whether to take precautions that cost c .⁴² If either or both parties to a pairwise interaction have taken precautions, both receive payoffs of 1; otherwise, both get 0. Thus, in a fully connected network in which p of the n players have taken precautions, each cautious player gets $n-1-c$ and each incautious player gets p , for a total payoff of $(2n-1-c)p - p^2$, so in an efficient network $p = \frac{1}{2}(2n-1-c)$. In such a network, the (pure-strategy) equilibrium level of p satisfies $p > n-1-c > p-1$; it is only efficient if $c = \frac{1}{2}$. Network formation is rather trivial: any incautious player wishes to link to any cautious player, and is indifferent as to whether he links to another incautious player. Any cautious player wishes to link to all other players, so the fully-connected network is (at least weakly) an equilibrium configuration. Clearly, this setup should be extended to a more realistic specification.

A7 The Hybrid Game

Description of the one-sided model

Consider a population of linked, but heterogeneous individuals, each of whom can choose between a high-trust and a low-trust strategy. Preferences are parametric – it should be clear how they can be derived from, for example, the coordination game (Section 2.2 in this paper). Define:

Variable	Definition
Δ_i	The payoff advantage to player i of using the high-trust strategy
ρ_i	Player i 's perceived risk-cost from using the high-trust strategy
h_i	The proportion of i 's neighbours using the high-trust strategy
θ_i	Player i 's idiosyncratic taste parameter, independently and identically distributed according to density $f(\theta)$ and cdf $F(\theta)$
v	The strength of the 'network externality' among high-trust players

We assume that players have the following evaluation of the high-trust option:

$$\Delta_i = \theta_i [1 - v + v h_i] - \rho_i$$

The 'taste' parameter θ_i measures the relative value of these benefits compared with the expected costs ρ_i , which include the cost of switching to the high-trust strategy.⁴³ θ is independently and identically distributed on a compact interval $[\theta^-, \theta^+]$.

The 'benefits' term includes both a fixed component (for example, the reduced transaction costs associated with trusting certified public channels) and a term that varies with the number of other high-trust individuals with whom player i interacts.

We assume a fully connected network involving a large population, so $h_i = h$, where h is the proportion of individuals employing the high-trust strategy. v measures the strength of the network interaction: if $v = 0$ there is no network externality, while $v = 1$ corresponds to the 'peer-to-peer' case where only network interactions matter. There is no asymmetry in risk cost assessments, so $\rho_i = \rho$ for all i .

Equilibrium behaviour

Nash equilibrium implies the set of high-trust players is $[\theta^*, \theta^+]$, where the cut-off value, θ^* , satisfies:

$$\text{Indifference: } \theta^* = \frac{\rho}{1 - v + v h},$$

The figure below plots these equations for $\rho = 1$, $v = 0.91$ and $\theta \sim N(0.87, 3.6)$.

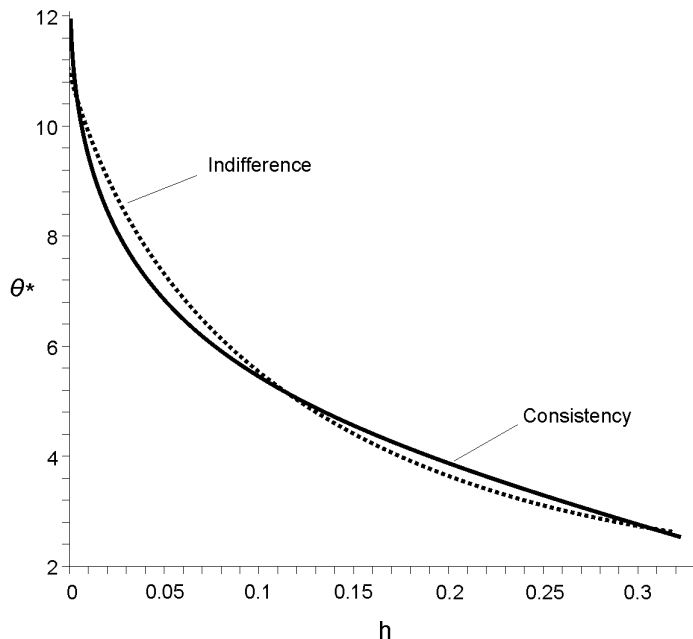


Figure 1 in the text plots the equilibrium values of h against ρ and v .

A two-sided version

Assuming fixed network effects, risk responds instantaneously to the prevalence of trusting behaviour: $dp/dt = \phi(h)$.

For predation or opportunism dp/dt is positive or negative if h is larger or smaller than a critical value h^* where expected returns to abuse of trust just balance expected costs. If network effects are small enough for trusting to respond continuously to risk ($n < n^*$), the model converges monotonically to $v=v^*$ and $\rho = [1-v+vh^*]F^{-1}(1-h^*)$. If network effects are 'too large' the system cycles (clockwise) in a hysteresis loop.

For reassurance dp/dt is positive or negative if h is above or below h^* . In this case, the system tends to one of the two corner solutions: high-trust ($h=1$ and $\rho=0$) or low-trust ($h=0$ and $\rho=1$). This result can also be obtained from a Bayesian model in which subjective risk estimates are adjusted according to a dynamic equation:

$$\rho_t = (1 - \lambda)\rho_{t-1} + \lambda h, \text{ or } \dot{\rho} = \lambda(h - \rho)$$

An incomplete-information model

Qualitatively similar results (roughly S-shaped time-paths for creation and erosion of trust) can be obtained from a simple incomplete information model where individuals become aware of current differential pay-offs to trusting behaviour by sampling the population. Dynamics depend critically on credibility and any bias in this information. One interesting feature is that S-shaped adoption paths can result from a uniform distribution of prior beliefs.⁴⁴ Using the above notation, the population of high-trust individuals at any given time is a representative sample of the distribution of tastes θ rather than an 'upper interval' of the form $[\theta^*, \theta^+]$.

A more explicitly dynamic story would model strategy choice as a two-armed bandit problem, in which the value of information is traded off against current payoff.

In the equilibrium model sketched above, average trust per capita is given by

$$\tau = \frac{\int_{\theta^*}^{\infty} [1 - \nu + \nu(1 - F(\theta^*))] \theta f(\theta) d\theta}{\int_{\theta^*}^{\infty} f(\theta) d\theta}$$

which for $\theta \sim N(\mu, \sigma^2)$ gives

$$\tau = [1 - \nu + \nu(1 - F(\theta^*))] \frac{\theta^* \sigma^2 f(\theta^*)}{1 - F(\theta^*)}$$

By contrast, in the epidemiological model with the same normal distribution, each type θ 's propensity to use the high-trust strategy is $\tau = \theta[1 - \nu - \nu h^*]$, so per-capita average trust is: $\tau = \mu[1 - \nu - \nu h^*]$.

We can compare these two expressions to track the response of trust to a secular decrease in perceived risk – equivalent to a fall in the critical taste parameter θ^* at which the individual is indifferent between high and low trust (cf Table 5).

Change in τ as θ^* falls	Equilibrium	Incomplete information
No network externalities ($\nu = 0$)	$\frac{\sigma^2 f(\theta^*)}{1 - F(\theta^*)}$	μ
Peer-to-peer ($\nu = 1$)	$\sigma^2 f(\theta^*)$	μh^*

A1.8 Network Dynamics in the Coordination Game

We use the game from Section 2.2 of this paper – if a fraction γ_H of a player's neighbours uses the high-trust strategy, she or he will also play high-trust if $\gamma_H > \frac{3-A}{8-A-B}$. A fully-linked group of $n+1$ players subject to random shocks will converge to high trust in the long run if $\frac{3-A}{8-A-B} < \frac{1}{2}$ – in other words if high-trust is better than low-trust against a random opponent ($\frac{5+A}{2} > \frac{B+3}{2}$).

Note the low trust equilibrium can be stable even though it is not efficient. Figure 2 in this paper shows the stable conventions as a function of the parameters A and B.

A1.9 Network Dynamics with Criminal Behaviour

We use the game from Section 3.3 in this paper – the stochastically stable conventions⁴⁵ for games with more than two strategies are those for which the 'resistance' of each convention to shocks that might 'tip' it towards another convention is greatest.

These shocks might be direct (copying one's neighbours) or indirect (responding to changes in partners' behaviour by adopting a third strategy). The following matrix shows the resistance⁴⁶ associated with each transition.

Transition	Direct resistance	Indirect resistance	Minimum
H→L	$\max\left\{\frac{5-B}{8-A-B}, \frac{3-B}{6-B}\right\}$	$\frac{5-B}{9-B}$ if $B \leq \frac{17}{5}$	See below
H→C	$\max\left\{\frac{2}{7}, \frac{B-3}{B-2}\right\}$	$\frac{2}{8-A}$ if $A < \frac{6(B-2)}{B-3}$	See below
L→H	$\frac{3-A}{8-A-B}$	N/A	$\frac{3-A}{8-A-B}$
L→C	$\frac{3}{4}$	$\frac{3}{6-B}$ if $A < \frac{24}{B-3}$	See below
C→H	$\max\left\{\frac{5}{7}, \frac{4}{9-B}\right\}$	N/A	$\max\left\{\frac{5}{7}, \frac{4}{9-B}\right\}$
C→L	$\frac{1}{4}$	N/A	$\frac{1}{4}$

The minimum resistance for the transition H→L is

$$\begin{aligned} & \frac{5-B}{8-A-B} && \text{if } \frac{6}{B-3} \leq A \leq -1 \text{ or } B > 3.4 \\ & \frac{3-B}{6-B} && \text{if } A \leq \frac{6}{B-3} \text{ and } B > -3 \\ & \frac{5-B}{9-B} && \text{if } B \leq 3.4 \text{ and either } B < -3 \text{ or } A > -1 \end{aligned}$$

The minimum resistance for the transition H→C is:

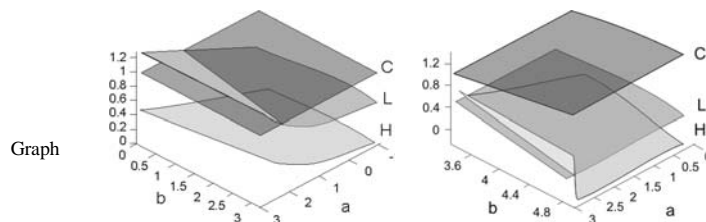
$$\begin{aligned} & \frac{2}{7} && \text{if } B \leq 3.4 \text{ and } A > 1 \\ & \frac{B-3}{B-2} && \text{if } A \geq 1 \text{ and } B \leq \frac{20-3A}{6-A} \\ & \frac{2}{8-A} && \text{Otherwise} \end{aligned}$$

The minimum resistance for the transition L→C is:

$$\begin{aligned} & \frac{3}{4} && \text{if } B < 2 \text{ or } A > \frac{24}{B-3} \\ & \frac{3}{6-B} && \text{otherwise.} \end{aligned}$$

To identify the stochastically stable convention, sum the resistances associated with all transitions towards each convention; the stable convention has the lowest sum:

Convention	Resistance	
	if $B < 3.4$	if $B \geq 3.4$
H (high-trust)	$\frac{32-13A-7B+AB}{(8-A-B)(9-B)}$	$\frac{61-12A-5B}{7(8-A-B)}$
L (low-trust)	$\frac{28-A-5B}{4(8-A-B)}$	$\frac{29-5B}{4(9-B)}$
C (crime)	$\frac{28}{29}$	$\frac{7B-18}{4(B-2)}$



A complete parametric map of the stable conventions is shown in Figure 3 in this paper.

NOTES

- 1 The author would like to acknowledge the support of the Department of Trade and Industry and stimulating conversations with members of the Foresight Cyber Trust and Crime Prevention project team (especially Miles Yarrington, Robin Mansell and Claire Craig), colleagues at RAND Europe (especially Maarten Botterman and Lorenzo Valeri) and the University of Warwick (especially Bhaskar Dutta and Myrna Wooders), seminar participants in Bled, Oxford, London, and Warwick and the helpful comments of anonymous referees. Of course, they bear no responsibility for any errors or omissions.
- 2 Testing the model is at best a matter for econometrics, but in practice many of the phenomena here cannot be measured and/or depend on individual and collective perceptions that may be hard to measure separately from the behaviour that the models attempt to describe.
- 3 The analyses are condensed (esp. in Section 2) for focus. More detailed treatment is available from the author on request.
- 4 E.g., trusting behaviour vs. trustworthiness, trust as a public or a private good; a matter of credibility or of delegation, etc.
- 5 In classical game theory, all 'players' interact with each other. More recently, these tools have been extended to take account of patterns of local interaction, see, e.g. Morris (2000); Jackson and Watts (2002).
- 6 These norms include billing later for services, acceptance of fiat currency, self-reporting in transactions, etc.
- 7 See e.g. David and Spence (2003).
- 8 The sociological literature has extensively developed such trust architectures, see, e.g., Eisenstadt and Roniger (1984). The computing literature, has further elaborated these constructs to assist communication in insecure environments.
- 9 See, Knight (1921).
- 10 e.g. Hardin (1991), Bowles and Gintis (2000) and Guerra and Zizzo (2002).
- 11 A partial exception is made in the 'crime game' examined in Sections 2.2 and 3.3.
- 12 See, e.g., Kandori et al. (1993).
- 13 See, e.g., Young (1998).
- 14 See, e.g., Jackson (2003).
- 15 Kasher and Rubinstein (1998) differentiate group identification based on individual affiliation, group inclusion and third-party labelling.
- 16 Asymmetry with regard to *roles* (e.g. buyer and seller) is perfectly compatible with symmetry across members of a society. If a strategy is described as a prescription for how a person should behave in a given situation and if any person might be in any position, symmetric equilibrium is much less restrictive (since the play between different roles can be asymmetric).
- 17 Including team working games like 'Stag Hunt.'
- 18 Like 'Battle of the Sexes'.
- 19 e.g. Bacharach. and Gambetta (2001), Bacharach et al. (2001), Bohnet et al. (2001), Bohnet and Zeckhauser (forthcoming 2004).
- 20 On this point, see esp. Bowles and Gintis (2000) and Ben-Ner and Putterman (2001).
- 21 These possibilities are sketched in the Appendix. In any case, the use of simplified representative or reduced form games has a long history: see e.g. Axelrod (1984), Morris (2000), Young (1993, 1998).
- 22 For instance, A might have different preferences than B believed (reputation – see Section 4.2.2), darker motives, a steeper discount factor, etc.
- 23 See e.g. Hintikka and Sandhu (1996) on the epistemic/semantic issues and Harsanyi (1973), Cho and Kreps (1987) on game-theoretic approaches. These approaches are combined to provide a framework for trust issues in Bacharach and Gambetta (2001).
- 24 We could equivalently assume that direct trust exposes a party to risk from all those trusted by his partner(s), but this would complicate the exposition without compromising the basic tension between efficiency and equilibrium.
- 25 Bacharach and Gambetta (2001) identify these as salient characteristics, though the underlying game and analysis are different.
- 26 If the network is not fully-connected, the possible trajectories may be more complex, since some players will face different circumstances.
- 27 See, e.g., Belleflamme and Bloch (2001), which characterizes *efficient* and *stable* market sharing networks in which firms agree not to compete with their 'partners'. Efficient networks in symmetric industries involve all firms having the same number of agreements; in stable networks completely-connected 'clusters' of a minimum size form – these clusters are typically too small to maximize industry profit and larger than would be socially optimal. Stable networks in asymmetric industries may be incomplete and fail to provide socially optimal levels of connection.
- 28 E.g., electronic input markets such as Covisint or MyAirline.com and an ever-expanding range of other B2B e-

- commerce markets. Many such markets have been set up by either the buyer or the seller side; the remarks above are concerned with their impact on trust among players on one side of the market only.
- 29 There is some evidence that it is relatively easy to attract customers to web sites, but much harder to retain them compared with offline channels.
- 30 By providing complete, accurate, reliable and durable information.
- 31 In other words, whether search results in greater satisfaction or a higher possibility of regret.
- 32 As demonstrated by recent 'phishing' attacks on online banking sites and transactions service providers such as Paypal, see <http://news.independent.co.uk/digital/features/story.jsp?story=473895> accessed 17 Apr. 04.
- 33 Effective consumer search is inhibited by information asymmetry, and the provision of comparative information inhibits treatment of difficult or doubtful cases, leading to a classic 'lemons market'.
- 34 Anderson (2003) cites the case of mobile telephones, where anti-cloning precautions led to increases in theft of 'legitimate' phones – and thus to an increase in physical risk accompanying financial risk.
- 35 At least, providing the 'hassle costs' of obtaining warranty performance are not too high. This is certainly an issue in the cyber world, where much of the 'protection' takes the form of customer service or advice which may take a long time to obtain and comes without further warranty for consequential damage.
- 36 Relevant articles include Shapiro (1982, 1983).
- 37 Calabresi (1970).
- 38 This is connected to the analysis of incomplete rationality and the insight that it can be Pareto-improving to commit to not using all relevant information, not auditing in every case where it is possible to do so and not retaining extensive records of past behaviour. See, e.g., Jensen and Meckling (1976) or Murphy (1999) for examples from financial economics.
- 39 See, e.g., Caillaud and Jehiel (1998), Fischer and Kakkar (2000), Geanakoplos, et. al. (1990), Jehiel and Moldovanu (1996), Jehiel et al. (1996) and Kubler and Schmedders (2003).
- 40 This is a simplified example of an optimal incentive contract – see e.g. Laffont and Tirole (1993).
- 41 We also assume $d(i,i) = 0$ and that $d(i,j)$ is infinite if i and j are unconnected.
- 42 Again, players could incur cost for each pairwise trust relationship, but this adds little to the indirect reliance analysis.
- 43 An equivalent formulation would assume quadratic utilities; treat r_i as the (subjective) variance of returns, and q_i as an inverse measure of risk aversion.
- 44 Jensen (1982) obtains a similar result in a technology diffusion model.
- 45 Following Young (1993).
- 46 Formally, the resistance is the proportion of 'mistakes' by neighbours needed to induce a player to make the indicated transition. For instance, if the current convention is C, a player will switch to L if at least $\frac{1}{4}$ of the other players deviate to L (direct resistance) or if at least $\frac{1}{(B-2)}$ of the other players deviate to H (indirect resistance; in this case the proportion playing H must not be larger than $\frac{4}{(9-B)}$ – since $B < 5$, the 'threshold' associated with the indirect transition is higher than that associated with the direct transition.

REFERENCES

- Anderson, R. (2003), 'Cryptography and Competition Policy: Issues with "Trusted Computing"', Cambridge University Working Paper, <http://www.ftp.cl.cam.ac.uk/ftp/users/rja14/tcpa.pdf> accessed 17 Apr. 04.
- Axelrod, R. (1984), *The Evolution of Cooperation*, New York: Basic Books.
- Bacharach, M. and D. Gambetta (2001), 'Trust in Signs' in K. Cook (ed.) *Trust in Society*, New York: Russell Sage Foundation, pp. 148-84.
- Bacharach, M., Guerra, G. and Zizzo, D. (2001), 'Is Trust Self-fulfilling: An Experimental Study', Oxford University Department of Economics Working Paper 76.
- Bailey, J. (1998), 'Intermediation and Electronic Markets: Aggregation and Pricing in Internet Commerce', Unpublished PhD Dissertation, Program in Technology, Management and Policy, Massachusetts Institute of Technology.
- Bakos, J. (1997), 'Reducing Buyer Search Costs: Implications for Electronic Marketplaces', *Management Science* 43(12): 1676-92.
- Belleflamme, P. and Bloch, F. (2001), 'Market Sharing Agreements and Collusive Networks', University of London Queen Mary College Working Paper 443, <http://www.econ.qmul.ac.uk/papers/docs/wp443.pdf> accessed 17 Apr. 04.
- Ben-Ner, A. and Putterman, L. (2001), 'Trusting and Trustworthiness', *Boston University Law Review* 81(3): 523-51.
- Bohnet, I., Frey, B. and Huck, S. (2001), 'More Order with Less Law: On Contract Enforcement, Trust and Crowding', *American Political Science Review* 95(1): 131-44.
- Bohnet, I. and Zeckhauser, R. (2004 forthcoming), 'Trust, Risk and Betrayal', *Journal of Economic Behavior and Organization*.
- Bowles, S. and Gintis, H. (2000), 'Optimal Parochialism: The Dynamics of Trust and Exclusion in Networks', Santa Fe Institute Working Paper.

- Brynjolfsson, E. and Smith, M. (2000), 'Frictionless Commerce? A Comparison of Internet and Conventional Retailers', *Management Science* 46(4): 563-85.
- Büschken, J. (2000), 'Reputation Networks and "Loose Linkages" Between Reputation and Quality', Katholischen Universität Eichstätt, Wirtschaftswissenschaftliche Fakultät Ingolstadt, Working Paper, <http://www.ku-eichstaett.de/Fakultaeten/WWF/Lehrstuehle/MKT/>
downloads accessed, 17 Apr. 04.
- Caillaud, B. and Jehiel, P. (1998), 'Collusion in Auctions with Externalities', *Rand Journal of Economics* 29(4): 680-702.
- Calabresi, G. (1970), *The Costs of Accidents: A Legal and Economic Analysis*, New Haven CT: Yale University Press.
- Cho, I. and Kreps, D. (1987), 'Signaling Games and Stable Equilibria', *Quarterly Journal of Economics* 102(2), 179-221.
- David, P. and Spence, M. (2003) 'Towards Institutional Infrastructures for e-Science', Oxford Internet Institute Research Report No. 2, http://www.oii.ox.ac.uk/resources/publications/OIISP_BailliolOII.pdf accessed 17 Apr. 04.
- Eisenstadt, S. and Roniger, L. (1984), *Patrons, Clients and Friends*, Cambridge: Cambridge University Press.
- Ellison, G. (1993), 'Learning, Local Interaction and Coordination', *Econometrica* 61(5): 1047-71.
- Fischer, E. and Kakkar, E. (2000), 'On the Evolution of Comparative Advantage', <http://econ.ohio-state.edu/efisher/evolution.pdf>, accessed 17 Apr. 04.
- Fukuyama, F. (1995), *Trust, the Social Virtues and the Creation of Prosperity*, New York: Free Press.
- Geanakoplos, J., Magill, M. Quinzii, M. and Dreze, J. (1990), 'Generic Inefficiency of Stock Market Equilibrium When Markets are Incomplete', *Journal of Mathematical Economics* 19: 113-51.
- Guerra, G. and Zizzo, D. (2002), 'Trust Responsiveness and Beliefs', Oxford University, Department of Economics Working Paper No 99.
- Hardin, R. (1991), 'Trusting Persons, Trusting Institutions', in R. Zeckhauser (ed.) *Strategy and Choice*, Cambridge MA: MIT Press, pp. 185-209.
- Harsanyi, J. (1973), 'Games With Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points', *International Journal of Game Theory* 2: 1-23.
- Hintikka, J. and Sandhu, G. (1996), 'Game-Theoretical Semantics', in J. van Benthem and A. ter Meulen (eds) *Handbook of Logic and Language*, Amsterdam: Elsevier, pp. 361-410.
- Hollis, M. (1998), *Trust Within Reason*, New York: Cambridge University Press.
- Jackson, M. (2003), 'The Stability and Efficiency of Economic and Social Networks' in S. Koray and M. Sertel (eds) *Advances in Economic Design*, Heidelberg: Springer-Verlag, pp. 319-362.
- Jackson, M. and Watts, A. (2002), 'On the Formation of Interaction Networks in Social Coordination Games', *Games and Economic Behaviour* 41: 265-91.
- Jehiel, P. and Moldovanu, B. (1996), 'Strategic Nonparticipation', *Rand Journal of Economics* 27(1): 84-98.
- Jehiel, P. Moldovanu, B. and Stacchetti, E. (1996), 'How (Not) to Sell Nuclear Weapons', *American Economic Review* 86(478): 814-29.
- Jensen, M. and Meckling, W. (1976), 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure', *Journal of Financial Economics* 3(3): 305-60.
- Jensen, R. (1982), 'Adoption and Diffusion of an Innovation of Uncertain Profitability', *Journal of Economic Theory* 27(1): 182-93.
- Kandori, M.G., Mailath, G. and Rob, R. (1993), 'Learning, Mutation, and Long Run Equilibria in Games', *Econometrica* 61: 29-56.

- Kasher, A. and Rubinstein, A. (1998), 'On the Question "Who is a J?" A Social Approach', Tel Aviv Foerder Institute for Economic Research and Sackler Institute for Economic Research Working Paper: 20/98.
- Katz, M. and Shapiro, C. (1994), 'Systems Competition and Networks Effects', *Journal of Economic Perspectives* 8: 93–115.
- Knight, F. H. (1921) *Risk, Uncertainty and Profit*, Boston MA: Riverside Press.
- Kubler, K. and Schmedders, K. (2003), 'Generic Inefficiency of Equilibria in the General Equilibrium Model with Incomplete Asset Markets and Infinite Time', *Economic Theory* 22 (12): 1-15.
- Laffont, J-J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge MA: MIT Press.
- Morris, S. (2000), 'Contagion', *The Review of Economic Studies* 67(1): 57-78.
- Murphy, K.J. (1999), 'Executive Compensation', in O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics*, Volume 3B, New York and Oxford: Elsevier Science, pp. 2485-63.
- Page, F., Wooders, M. and Kamat, S. (2002), 'Networks and Farsighted Stability', University of Warwick Economics Research Papers No. 621.
- Pollitt, M. (2001), 'The Economics of Trust, Norms and Networks', Judge Institute of Management Working Paper, Cambridge University, <http://www.econ.cam.ac.uk/electricity/people/pollitt/economictrust.pdf> accessed 17 Apr. 04.
- Putnam, R. (2000), *Bowling Alone - The Collapse and Revival of American Community*, New York: Simon & Schuster.
- Rifkin, J. (2000), *The Age of Access*, New York: JP Tarcher/Putnam.
- Shapiro C. (1982), 'Consumer Information, Product Quality, and Seller Reputation', *Bell Journal of Economics* 13(1): 20-35.
- Shapiro C. (1983) 'Premiums for High Quality Products as Rents to Reputation', *Quarterly Journal of Economics* 98: 659-680.
- Watts, D. (1999), *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton NJ: Princeton University Press.
- Young, P. (1993), 'The Evolution of Conventions', *Econometrica* 61(1): 57-84.
- Young, P. (1998), *Individual Strategy and Social Structure*, Princeton NJ: Princeton University Press.