An Introduction to the Literature on Online Reputation Systems for the MMAPPS project

James Howison

August 18, 2003

1 Introduction

This document provides an introduction to the research literature on online reputation systems. It has been written with a view to providing background and context for MMAPPS system design plans and is therefore focused on distributed systems. There is a companion annotated bibliography of summary and reading notes for many of the articles discussed in this document in the accompanying References.

The latex source of this document is available[1].

1.1 What are online reputation systems for?

Online Reputation systems are looked to as possible solutions to these problems in distributed systems, motivating co-operation, providing recommendations and as a distributed authorisation mechanism. It is the first of these, motivating co-operation, that seems the most general and relevant to the MMAPPS project.

Motivating co-operation

Co-operation is of particular interest in distributed systems because there is no central authority capable of imposing monitoring and or sanctions on the behaviour of peers. Networks services, such as search, operate through the co-operation of the peers, not the the imposition of hierarchical authority. Engendering co-operation in distributed systems in critical to their success. Reputation systems are suggested as being valuable in motivating co-operation between individuals and in groups.

Motivating co-operation has a strong history in research literature. There is an active strand of evolutionary biology that has studied altruism, seeming "unselfish behaviour", in the group behaviour of primates and humans (Trivers, 1971; Alexander, 1987; Gintis et al., 2001; Nowak and Sigmund, 1998; Sugden, 1986; Pollack and Dugatkin, 1992; Fehr and Gächter, 2002). These authors have sought to understand, through simulations and analytical game theory, the reasons behind

[1] http://www.mmapps.org/private/material/papers/RepLitIntro/sources/

this co-operation. Their motivator is "evolutionary fitness" which is usually expressed as the ability to pass on their genes (or strategies) to future generations.

Co-operation has also been studied in Economics, which has drawn on some of the literature cited above, but where it is typically modelled as a variant of the Prisoner's dilemma[2] and iterated or repeated dilemmas (Granovetter, 1985; Milgrom and Roberts, 1982; Fudenburg and Tirole, 1991; Marinoff, 1992). The Prisoner's dilemma is conceived of as being analogous to a situation faced in a simple market economy between a buyer and a seller without legal enforcement of the exchange contract. This work proceeds through game theoretic analysis of payoffs and strategies with simulations being employed to assess the success of strategies in the repeated games. As these papers typically understand their task as modelling a market economy, the metric for successful strategies is typically understood as accumulation of individual utility.

The economic strand of work has developed into proposals and metrics for economic systems, such as auction designs and systems for choosing public goods. This work is best known as "Mechanism Design" which develops from the game theoretic analysis systems designed to enforce co-operation through strategyproofness and/or incentive compatibility. The mechanism is designed so that it is in the best individual interest of the players to co-operate and in the best overall interest of the group to co-operate. A useful introduction to this relatively complex work is provided by Jackson (2000).

The papers explicitly proposing reputation systems for online distributed systems engage with this research literature in inconsistent ways. It is often unclear whether the model and solution are an application of elements of this theory or whether the situation is novel. Understanding the implications of the distributed environment for the solutions and proofs offered in the literature is an open research task. This task will cast more light onto the proposals considered below.

This author began by reading the system proposals and moved backwards into the underlying literature—a process which often lead to a reconsideration of the proposal papers and their strengths and weaknesses.

This strand of work has been applied to Grid computing systems by Azzedin and Mahewswaran (2002). Readers should be aware that there is a parallel strand of literature on co-operation which focuses on token payments systems. The similarities between token based systems and reputation based systems are discussed by Moreton and Twigg (2003).

Recommender systems

Recommender Systems provide information on the opinion of a user on the quality of a (usually creative) work. The best known examples are the recommender systems associated with Amazon ("People that bought this book also bought these books"). Another well known and developed system is MovieLens, part of the GroupLens project.

The Gnutella focused systems designs (Kamvar et al., 2003; Cornelli et al., 2002) target content quality through a provider reputation for the provider and are thus hybrids.

[2] http://www.wikipedia.org/w/wiki.phtml?search=prisoner%27s+dilemma&go=Go

Distributed authentication

Authentication (Web of Trust). Reputation has also been proposed as an element of systems aiming to approximate the capabilities of the Public Key Infrastructure ("PKI") in a distributed environment. PKI systems such as SSL certificates are issued and signed by certificate authorities in an ascending hierarchy. By contrast the PGP Web of Trust does not rely solely on hierarchical trust lines but can assess validity through the signatures of known peers. Eschenauer et al. (2003) writes of the possible uses of reputation for Authentication.

It may be that this is a sub-set of the more general reputation concepts in which the only attribute to be established is identity, rather than the more usual concepts of trustworthiness etc. For MMAPPS the more interesting questions are how one will trust these identities.

1.2 Goals of reputation research

The research work into online reputation systems is focused on the theory of co-operation and the design of systems. In particular these features are examined:

- Design—specifying assumptions, environment and behavioural model utilised.
- Performance—How does the proposed system perform, does it motivate co-operation? Under what assumptions? This is typically assessed through a simulation involving a set of both mixed and pure strategies. There are a number of papers that attempt to derive dominant or evolutionary stable strategies and prove their performance that way. Performance of strategies is measured through a success metric. The evolutionarily focused simulations utilise a metric termed "evolutionary fitness" in which accumulation of payoffs increases the probability of survival through to the next generation. The more economically focused simulations utilise the accumulation of utility during one lifetime as their metric.
- Efficiency—Does the proposed model lead to economic efficiency. This is typically measured in aggregate utility - a simple summation of the utilities gained by participants in the games. This notion can be seen in the assessment that if both co-operate the sum of their achieved utilities is the highest that could have been extracted from the game.

  It should be noted that there is a strong debate over the correctness of this measure of economic efficiency, which draws on the notion of additive utility. Some Economists prefer to utilise ordinal utility[3], which only requires an ordering of individual preferences and are not directly comparable across individuals. The assessment of efficiency when utilising this measure is not straightforward - additive and aggregate utility appear to have been chosen for their simplicity on an issue which is tangental to the main research and design goals.

  However even when additive utilities are used aggregation may not be the only goal for efficiency in the system. Although not typically considered in the papers examined social fairness and distributional considerations can be used in the assessment of the efficiency of a system.

  Note that the assessment of efficiency typically applies to the market in which a reputation system is being applied—not directly to the reputation system itself. That is to say that the

[3] http://cepa.newschool.edu/het/essays/paretian/paretosocial.htm

reputation system is an adjunct designed to increase the efficiency of the system—rather than a goal in itself.

- Computational Efficiency—A growing field of research which stems from Mechanism Design is attempting to assess the computation efficiency of mechanisms. Most Mechanism design assumes that the calculations required would be performed by a trusted centralised agency (in public choice this is conceived as a disinterested state). Within a peer-to-peer system this is clearly problematic and so research has focused on iterated and distributed algorithms. The Economics & Computer Science Research Group[4] at Harvard is central to this research. For an introduction to this work see Parkes (2001).

1.3 Reputation in MMAPPS

Reputation is of interest in the MMAPPS project as a support mechanism for the peer-to-peer interactions which form the basis of the MMAPPS framework. In particular reputation systems are relevant in three ways.

- Reputation for Trustworthiness—A system to provide assurance that the description of the interactions contained in the accounting system is reliable. The system must deal with the problematic distributed environment (as described in Issues for distributed reputation systems in Section 3.3) and may achieve its goals by methods which differ in their attitude to the truth of accounting reports.
  - Motivating truth. co-operation (truthful reporting) a preferred strategy.
  - Punishing Falsehood. rewarding an auditing activity such that the threat of auditing means that peers prefer truthful reporting
  - Ignoring truth/falsehood in particular transactions but rewarding truthful parties in the long-run. The proposal is to punish both parties in a disputed transaction such that repeatedly dishonest parties will be punished proportionally more than truthful parties.
- Reputation for Quality—A system to provide information on the subjective quality of the services provided by peers. This system may operate as a service utilising the framework. It seems to be similar to a recommender system and is an assistance to peers seeking services whose subjective quality they will appreciate.
- Reputation focused accounting systems—This is an open issue whose exploration is part of what MMAPPS is all about.

The concepts of Reputation for trustworthiness, focused on motivating co-operation and incentive compatibility, are closest to the literature introduced in this document and the accompanying annotated bibliography.

It is an open question whether these systems can be safely implemented in an entirely peer-to-peer fashion or whether centralisation of this function is desirable. If centralisation is desirable then these roles represent an exploitation opportunity for the MMAPPS partners.

[4] http://www.eecs.harvard.edu/econcs/

1.4 Concepts

There are a number of concepts that are key to engaging with the literature on reputation and reputation systems. Mui et al. (2002) develops a useful taxonomy of reputation types.

Reputation as a Signal and a Sanction

Reputation can act as a signal or as both a signal and a sanction. This is mentioned briefly by Resnick and Dellarocas (2003) but was a major theme at the MIT Online Reputation Symposium in May 2003[5] which was blogged[6] by the author. Interestingly this distinction is not discussed in the rest of the literature examined, although it is implied by some modelling decisions.

Reputation as a sanction—If agents are able to alter their behaviour when playing a game (i.e. can choose to co-operate or to provide higher quality service) then reputation can act as a sanction. It is a sanction because it motivates co-operation in order to build 1st order reputations. This is either through punishment of a negative reputation report or the benefit of a reputation report. But this raises the question of in what way is a good reputation a benefit and a negative reputation a problem? This is due to the role of reputation as a signal.

Reputation as a signal—Reputation acts as a signal in that agents choosing amongst a number of possible partners will be more likely to interact with agents with higher reputations for co-operation (this is true whether nor not they intend to co-operate or defect). It is via this signalling mechanism that reputation benefits a peer.

This feature of reputation is used in simulations as the probability of interaction. For example in Jurca and Faltings (2003), "two agents (A and B) are randomly selected to play the game ... After knowing the reputation of their partner, agents can decide whether or not to play the game".

Only in situations in which agents cannot alter their behaviour does reputation act only as a signal ('don't use that route - there is a 56k modem as a bottleneck'). This is the case for reputation that adheres not to people but to items, such as songs.

It can be seen that the role of reputation systems in altering behaviour, such as promoting truthful reporting of accounting information, relies on the ability of the reputation score to provide benefits through signalling such that the motivation to achieve good reputation scores acts as a sanction on behaviour.

It is significant, then, that the simulations utilized in the literature often assume that the partners are of a fixed quality; "We assume that the quality of a seller does not vary over time" (Miller et al., 2003) and "The sellers (or products) ... have an innate predetermined quality level, and play no active role in our model" (Miller et al., 2003, p 8). Jurca and Faltings (2003) adapts this only slightly by including a "mood" component based on the preceding moves (perhaps not as arbitrary as it seems - the "mood " might better be expressed as the agent's assessment of the nature of the market it is transacting in).

[5] http://www.si.umich.edu/~presnick/reputation/symposium/
[6] http://www.freelancepropaganda.com/archives/000031.html

Reputation as contextual information

It is normal in the system design literature to emphasise that a node's reputation is for a particular thing—which is to say that the reputation score has a particular context. One has a good reputation for a specific thing—be that always paying bills, adequately completing a given computational task etc.

However some papers go further and analyse this aspect of reputation in more detail. Sabater and Sierra (2000) argue that "reputation is compositional, the overall opinion on an entity is obtained as a result of the combination of different pieces of information". They call this the "ontological dimension" of reputation. Mui et al. (2002) states this boldy, "Reputation is clearly a context-dependent quantity. For example, one's reputation as a computer scientist should have no influence on his or her reputation as cook".

Contextual reputation allows us to conceive of having a reputation for providing accurate information on the reputation of others—what one might call second order reputation. A reputation for providing information on the second order reputation of others would be third order reputation, and so on. Jurca and Faltings (2003), amongst others, propose that one should buy reputation information from Reputation agents - and that one should decide which agents on the basis of their reputations for providing good reputation information, that is on their second order reputation. This distinction will become important when we examine the notion of influence in Section 3.1.1.

However, Gintis et al. (2001) suggests that this strict separation of reputation into different components and contexts might disguse the manner in which reputation works. Gintis et al. argue that altruistic behaviour is motivated because, despite its cost to the individual, it provides a signal as to the quality of that individual. These acts of generosity indicate to others that interaction with that individual in other contexts will be beneficial. - their term is that the "costly signalling" of altruism indicates an individual of "high quality". This quality is defined as, "genetic or phenotypic attributes that are difficult for others to assess directly, yet have important effects on the payoffs from social interactions with the signaler".

Thus it is often assumed that reputation is highly contextual, 'if you act well in one context I know only about that context and not about others', this may not in fact be so clear. One's actions in one sphere may be used to signal useful information in other contexts. This insight will become important when we examine re-cursiveness in reputation systems in Section 3.1.

2 Literature Review

The attached Annotated Bibliography provides further information on these papers but this section attempts to outline thematically their contents. The Annotated Bibliography also includes papers and comments not referenced in this section.

2.1 Analytical Studies of motivation

These papers address the theoretical issues in studying altruism or co-operation by humans.

Trivers (1971) developed a theory of reciprocal altruism which was generalized through the work of Axelrod (1984). This provides the classic introduction to the question of altruism by examining the

iterated prisoner's dilemma game and the results of a strategy tournament. This book summarized Axelrods previous years of research which established "Tit for Tat" as the most successful strategy. This strategy operates in the environment where there are repeated games with a good chance of meeting the same partner again. The strategy involves co-operating on the first turn, punishing a defection with a defection and then returning to co-operation.

There is a useful low-jargon introduction to this research available online[7].

Nowak and Sigmund (1998) further developed research into co-operation with their theory of indirect altruism. It is the closest analogy to the underlying assumptions behind most propositions for reputation systems.

Their basic proposition is that,

> Discriminators provide help [co-operate] to those individuals who have provided help. Even if the help is never returned by the beneficiary, or by individuals who in turn have been helped by the beneficiary, discriminating altruism [can be a successful evolutionary stable strategy] ... In this case, one does not expect a return from the recipient (as with direct reciprocity), but from someone else ... a donor provides help if the recipient is likely to help others' (which is usually decided on the basis of experience, i.e. according to whether the potential recipient has helped others in the past).

The reputation reflects the memory of those who have provided, and are therefore assumed to provide help in the future, to others.

Gintis et al. (2001) developed the theory of "costly signalling", a rather opaque name for a subtle concept. Gintis et al. argue that altruistic behaviour is motivated because, despite its cost to the individual, it provides a signal as to the more general quality of that individual. These acts of generosity indicate to others that interaction with that individual in other contexts will be beneficial. - their term is that the "costsly signalling" of altruism indicates an individual of "high quality". This quality is defined as, "genetic or phenotypic attributes that are difficult for others to assess directly, yet have important effects on the payoffs from social interactions with the signaler". Thus it is the "fuzziness" of reputation that motivates seemingly altruistic actions in some circumstances.

2.2 Experimental Studies of Motivation

There are two strands of related research which place themselves in the line of research described above but which approach the question not through analysis of game theory but through experiments. Their explanatory frameworks utilise characteristics which are innate to human beings, rather than a formalised rational payoff structure.

Fehr and Gächter (2002) develop a theory of "Altruistic punishment". They argue that there is a 'tendency' in humans to punish those that deviate from social norms. Clearly this is a theory that is only accessible through experiment rather than the analytical proofs which dominate the theories above. Their experiments where designed to exclude any of the mechanisms above, and to make punishment a costly activity—yet it still occurred. While they take care to place their work in the line of authority above their explanatory framework differs in that it relies on exogenous motivations rather than endogenous payoffs. Fehr is quoted in Grimes (2003) arguing, "I think

[7] http://www.abc.net.au/science/slab/tittat/story.htm

that trust has an emotional component and a cognitive component. It is important to understand both".

Grimes (2003) provides an accessible summary of the work that styles itself as a new field of 'Neuroeconomics'. Similar to Fehr's work this proceeds through experiment and understands its task as understanding actual human behaviour rather than providing analytical proofs. Zak et al. (2003) describes experiments which indicates higher rates of the secretion of oxytocin during episodes of co-operation. It is argued that oxytocin is a hormone associated with pleasant episodes and thus that trust has its own reward. Again the difference in explanatory framework from the analytical studies is clear.

2.3 Free Riding literature

The questions of motivation researched above are often summarized in the concept of the Tradgedy of the Commons, or the free-rider problem. There is a group of literature that takes this situation as its starting point.

Worthy of special mention is Pettit (1986) which outlines the free-rider problem in detail while introducing the concept of the "foul dealer". He summarizes it thus, "The free rider seeks to benefit by the efforts of others, the foul dealer to benefit at their expense"(Pettit, 1986, p 374) which is to say that a foul dealer does not only not co-operate but is able to take advantage of the co-operators in such a way that the co-operators are worse off then if they had not co-operated at all. Disarmament is taken as the paradigmatic example; while if all disarm each is better off, if any one nation does not and can invade any of the others then, in the presence of just one armed nation, each of the disarmers is worse off than if none had not disarmed.

2.4 Reputation System Proposals

The papers typically proceed through a short background piece, followed by a description of a trust model, then an experimental simulation and an interpretation of the results. Despite the use of the phrase "incentive compatible" there are few attempts to prove strategyproofness or incentive compatibility in the sense that it is employed in analytical mechanism theory.

Crucial issues to consider in analysing these systems include:

1. The generality of the approach
2. The adequacy of the simulations to consider all strategies.
3. The adequacy of the limits revealed by the simulations
4. Consideration of the role of layers such as the data storage layers to become home for attacks on the systems—There is some work emerging on the use of distributed hash tables (DHTs) and other distributed data structures to avoid these attacks. Aberer and Despotovic (2001) gives the greatest consideration to this issue.
5. Centralised vs decentralised systems
6. Whether the system merely identifies problematic players or content (signals), or whether it motivates co-operation (sanctions).

Mui et al. (2002) in "Notions of reputation in multi-agents systems: A Review" provides a useful bridge between the theorectical literature outlined above and the systems literature outlined below. They also provide a useful literature review and a hierarchy of reputation concepts. They also design a simulation framework to test a variety of reputation concepts—their simulation is thus not of a specific proposed system but of the functioning of various concepts of reputation.

Thus they test Encounter-derived individual reputation, Observed individual reputation ("Each agent designates 10 random agents in the environment as being observed"), Group-derived reputation (reputation imputed from Group membership) and Propagated reputation.

These concepts are tested through a evolutionary simulation in which there are four strategies, Always cooperate, Always Defect, TFT and their test strategy, Reputation Tit for Tat (RTFT) which is identical to TFT except that rather than co-operating on the first turn the reputation (judged in one of the four ways described above) is used to predict the partners actions (if likely co-operate then co-operate, if likely defect then defect). Their analysis supports the use of the notion of propagated reputation which significantly out-performed the other reputation notions. It should be noted that the types are fixed here—the reputation system allows the agents to 'learn' the type of their perspective partner. Thus reputation here is signalling, not sanctioning.

Sabater and Sierra (2000) contribute the notion of reputation as a composite of reputation for constituent parts. e.g. the reputation for being a good travel agent is composed of reputation for access to tickets and advice. They also utilise group reputations (both about groups and reputation information from groups). They provide formulas for calculating reputation scores for both ontologically simple and complex situations. They appear to assume access to a correct and centralized information store. Their simulation is focused on an interesting putative function of a reputation system—they measure the speed with which their system finds an agent who "behaves reliably until he reaches a high reputation value and then starts committing fraud". They argue that their system will discover such activity quickly.

(Aberer and Despotovic, 2001) present a proposal for a system capable of providing reliable reputation information, "In this paper we present an approach that addresses the problem of reputation-based trust management at both the data management and the semantic level". They consider their semantic proposal within the limitations of a p2p situation - they intend to store only complaints, not positive votes and utilize the 'retaliation' situation (in which the cheater complains about the partner it cheated). In this sense it is very similar to the proposed MMAPPS reputation system.

They store the complaints in a distributed data structure called a P-Grid which they claim will be robust and redundant enough to handle attempts to manipulate the system. They also utilise a concept of propagated reputation to check the likely veracity of the witnesses—despite stating that reputation is strictly contextual they appear to use the same measure of trustworthiness for veracity as they do for judging trustworthiness in transactions.

They do not consider the motivation or incentive compatibility issue—however if an agent were to lie and say that their partner had cheated when they had not, they too would be impacted by a complaint. It seems even less likely that there could be an incentive to not file a complaint when the partner had cheated—however there might be an incentive if the agent wished to keep his number of complaints below a threshold and there would no additional gain for the individual (assuming that the cheating agent filed a complaint to 'cover his trail').

Their simulation again assumes types for the agents, and the task is to identify those agents that are likely to cheating. They consider both cheating at the transaction and at the data storage layer—

agents cheat with the same likelihood whether they are transacting or reporting stored information. They claim that their simulation results indicate that the system does a good, but not perfect job of locating cheaters and conclude, "They show that a dectection of cheating agents is in fact possible in a pure peer-to-peer environment based on a fully decentralized method". Aberer and Despotovic are based at EPFL in Lausanne, Switzerland.

Side-market approaches

These papers utilise a side market for reputation information—the reputation reports are bought and sold to reputation agents of some form.

Miller et al. (2003) draw directly on the Mechanism Design literature and consider their problem to include motivating subjective effort in evaluation. This subjective effort in evaluation is not required in a Prisoner's Dilemma modelled situation - it is assumed to be costless to rate a partner as 'co-operate' or 'defect'. Consistent with the literature on Mechanism design their system uses a 'centre' to process complaints and to decide on the appropriate rewards and payments—in the process balancing the budget.

This mechanism has properties (such as simultaneous announcement) which may render it not useful for a system such as MMAPPS—however the author of this paper acknowledges his incomplete understanding of Mechanism Theory.

Jurca and Faltings (2003) proposes a system of Reputation-agents ("R-agents") that buy and sell reputation information on prospective partners. Players choose R-agents on the basis of their personal (i.e. non-propagated) experiences with the success of that particular R-agent. They create a side-market and restrict is such that there is no exchange between the currency used to pay partners and to pay R-agents. Yet this is problematic because there are additional prisoner's dilemma's in dealing with the R-agents—why would a further R-R-agent market not increase the efficiency of this market?

They claim to show that if an agent is only influenced by an a-priori type, then there is no payment function that can stimulate truthful reporting.

The price paid for the reports is structured such that truthful reports are paid above false reports. Truthfulness is judged as statistical similarity - in Jurca and Faltings (2003) it is judged truthful if it is the same as the last report about that agent. They therefore have the property that, if a significant proportion of the agents are lying, then the definition of truth is inverted. However this proportion is quite high—one would not expect that proportion of people to cheat. In their simulation the figure proves to be about 40%. The problems of agents systems relying on human behavioural insights is addressed in Section 3.1.

They also propose a useful way to register complaints and positive votes—prior to the transaction a partner provides a signed $rep_{my} + +$ vote and a signed $rep_{my} - -$ vote. After the transaction the partner chooses which to submit. Thus one's reputation can only be affected by oneself. Jurca and Faltings claim to be implementing their system on the EU 5th Framework funded Agentcities system[8].

[8] http://www.agentcities.org/

Non-database approaches

All the other system proposals considered in this paper approach the reputation problem as one which involves creating a database to store and access reputation information. The questions considered are, how to motivate the truthful insertion of informaiton into the database, how to store the database in an attack resistant manner and how to access and use the reputation information in making a decision.

Brainov and Sandholm (2002) does not propose to create a database of reputation information. Instead, Brainov and Sandholm propose that a seller can be motivated to truthfully reveal his/her quality to a prospective buyer. This is a fairly non-intuitive proposition which relies on the buyer knowing the marginal cost function of the seller. They summarize thus, "Instead of relying on a third party [or a database] for providing information, we design a class of market mechanisms in which agents reveal truthfully their levels of trustworthiness". It is an interesting paper that approaches the question from a fresh perspective but it appears to be limited in application to fairly narrow class of situations.

Gnutella quality focused proposals

Cornelli et al. (2002) have designed an addition to the Gnutella protocol that allows peers to be rated on the basis of the quality of the files they have provided. Their paper considers the provision of low-quality files on Gnutella as an absense of co-operation. They attempt to deal with the problem of Gnutella being based on throwaway identities by assigning zero reputation scores to new identities. They considered propagated reputation under the rubric of "enhanced polling". Their protocol for discovering reputation information is similar to the gnutella search protocol. They acknowledge the "overload the most reputable source" as a problem and suggest a domain specific response where the reputable sources serve MD5 hashes which then allow agents to pre-judge the quality of the files accessed from other clients (although why non-reputable agents wouldn't spoof the MD5 signature isn't considered ...)

Kamvar et al. (2003) in "EignenRep: Reputation Management in P2P Networks" Kamar, Schlosser and Garcia-Molina, from Stanford, present a reputation system focused on preserving the quality of files in a file sharing network. Similarly to Cornelli et al. (2002) their system uses quality votes to assign reputation to peers (using an opaque identifier - here they suggest Gnutella username). The reputation scores are then used to probabilistically determine which peers, from a set of responses to a query, a peer will download a file from. They adjust this probability to avoid the network converging on a single most trusted peer.

The system suggests the utilisation of a Distributed Hash Table for data storage and presents an algorithm for distributing the calculation of global trust scores for each peer. They do not attempt strategyproofness but consider a number of attacks which operate via the data storage layer and through manipulation of the semantic votes themselves. Their simulation, modelled on real world characteristics of the Gnutella network, compares a network without a trust system with their system under four situations of attack by malicious peers and groups of malicious peers. They do not simulate attacks on the Data storage layer, having attempted (but not proved) to dispel these in the design of the DHT.

One problematic element with both the 'quality files on Gnutella' papers, (Kamvar et al., 2003; Cornelli et al., 2002) is that there is little consideration of the motivational effects of the reputation—it

is not clear how a node is rewarded for having a good reputation—in fact nodes with high reputations will have more of their bandwidth used as users swarm to them. (Kamvar et al., 2003) merely suggests, "reputatable peers may be rewarded with increased connectivity to other reputable peers, or greater bandwidth".

2.5 Empirical Studies

Tadelis (1999)

All the eBay studies quoted in Resnick and Dellarocas (2003).

Resnick 1999 - They point out in a footnote to this sentence, however, that "Despite the rational incentive to free-ride, provision of feedback at eBay is quite common, occurring in more than 50% of transactions ...". Overwhelmingly positive - with a great deal of cross-citation.

3 Open research questions

3.1 Motivating sharing of reputation reports

Why would individuals contribute reputation reports to a globally accessible database? Furthermore why would they do so truthfully? These two questions are crucial in understanding reputation systems.

Initially it appears that there are, in fact, counter-incentives to doing so. Jurca and Faltings (2003) outlines this, "It is however not at all clear that it is in the best interest of an agent to truthfully report reputation information". They cite three reasons, the most important being that "it provides a competitive advantage to others" by assisting them to better predict defections and choose partners. Furthermore they point out that providing a positive rating (and thus increasing the others reputation) can decrease the value of ones own reputation, while providing a negative rating would boost ones own reputation, thus setting up a biased motivation for reporting.

Miller et al. (2003) also usefully outlines the motivation problem, "Evaluations may be underprovided, since providing feedback about a completed transaction requires some effort, and to improve the evaluation takes more, yet the information only benefits other participants". This problem is particularly obvious in the Gnutella quality papers, such as Cornelli et al. (2002), in which they acknowledge that those sites with a high reputation will be swamped by agents attempting to download—activity which will reduce the reporting agents' ability to use the network, thereby creating a disincentive for agents to share the location of high-quality providers with other agents [9].

These seemingly problematic incentives and the difficulties in understanding motivations, discussed below, are, however, belied by the empirical evidence of copious and seemingly keen provision of feedback on systems such as eBay and epinions Resnick and Dellarocas (2003).

[9]The strategies suggested to avoid this situation either involve some form of randomisation, which is essentially a controlled poisoning of the reputation information. Another attempt to avid this problem, using the reputable peers to provide MD5 hashes of the what quality is, merely shifts the problem such that the swamping will not be of agents download songs, but MD5 hashes of songs, which while providing some relief is obviously not going to scale

However there are a number of papers that would criticise the assumption that motivation is a crucial problem. These papers tend to rely on observed human behavioural characteristics. Fehr and Gächter (2002) observes a tendency to punish in spite of a lack of incentives to do so and proposes an explanation that there is a 'built-in' human pay-off in punishing through the act of punishment. Similarly the work discussed in Grimes (2003); Zak et al. (2003) it is argued that co-operation produces biological payoffs which would motivate the provision of reputation information as an altruistic act. Both these tendencies would counteract the disincentives discussed above.

Yet, It is far from clear how to quantify this pay-off, nor whether it is reliable when its costs are recognised. Furthermore in agents systems, at least, these human behavioural characteristics might not be operable—it might be that not all agents will represent humans or even be programmed by humans. Even if all agents where programmed by humans,relying on that human to programme responses that act at an emotional or biophysical level and might not even be consciously recognised, seems ill-advised.

The more usual strategy in the papers is to recognise these problems of motivation and to design the system to ensure that the payoffs are structured such that truthful reporting is the best rewarded strategy.

The most common strategy, followed in Jurca and Faltings (2003) and Miller et al. (2003) is the creation of side-markets for reputation information. These proposals are discussed in more detail at Section 2.4. By rewarding reputation reports with currency the system creates rational reasons to provide reports, and by rewarding truthful (in terms of statistical similarity) reports above others the system creates rational reasons for truth-telling.

Yet these strategies must create new markets which are surely inhabited by prisoner's dilemmas. In the case of Jurca and Faltings (2003) these are between the agents and the R-agents [10] and for Miller et al. (2003) between the agents and the "centre". It appears that the problem is therefore recursive—any system that assists the first market would, in turn, be similarly assisted by its own reputation system.

One possibility in resolving this issue is suggested by the work of Gintis et al. (2001). His argument, that altruism in one context signals 'quality' that is rewarded by increased opportunities in other contexts might provide a way out of the recursion. If the system is such that the provision of useful (truthful) reputation reports makes agents more likely to choose to undertake transactions with the reporting agent then the reporting agent would benefit for his reports through a greater number of profitable transactions.

3.1.1 Influence - 2nd order reputation

As discussed in Section 1.4 there is a gap in the literature in that few studies allow the quality of the seller, its liklihood of defection, to vary. And none have, as yet, addressed the situation in which the behaviour of the seller is altered depending on the reputation of the buyer he is currently dealing with. Investigating this area may prove fruitful.

[10] Jurca and Faltings (2003) attempts to resolve this by utilising a reputation system for the R-market, but limiting it to the personal experiences of an agent with the R-agents—yet if the propagated reputation represented by the R-agents makes the main market more efficient then why could such a system make the R-market more efficient. This special R-market would also require a limit on the number of R-Agents in order for personal experiences of agents to be of use to them in choosing between R-agents. Their solution to the recursion is not convincing

If an agent has a strong reputation for providing truthful and therefore useful reputation reports on other agents then it seems clear that this agent would be "listened to" (or weighted) by a larger number of other agents—each seeking the best reports. Thus we can say that such an agent is *influential*. The agent has influence because a report from that agent carries more weight or reaches more people than the norm, and thus has the potential to impact the subject of the report in a greater manner, negatively or positively.

Thus an influential agent is one that has a strong second-order reputation.

But what what motivate an agent to provide truthful reports and therefore earn a strong second-order reputation? This is only clear if we assume that agents can adjust their propensity to cheat depending on their current partner. If an agent, A, encounters a partner who has a strong second-order reputation, Inf, then it seem clear that that A will be less likely to defect when playing Inf than when playing an agent without influence. This is because a complaint from Inf would be more damaging to A's future prospects. Thus Inf is motivated to provide truthful reputation reports because it provides a direct benefit in terms of increased co-operation.

In this sense the provision of the reputation reports is a "costly signal" which could be interpreted as an indication that that agent is of "high quality" and therefore less likely to defect. Note that if one considers reputation to be strictly contextual then this does not make sense—only if reputation is indicative of a "higher quality" does it makes sense [11].

An agent, $UnInf$ with a low second-order reputation, is considering a transaction with agent A with an average second-order reputation. $UnInf$ consults Inf about the reputation of A—will Inf be able to predict the action of A when confronted by $UnInf$ remembering that it will be different than the action of A when confronted by Inf. Inf would need to report the likely action of A when confronted by someone of low influence. This is because Inf is not learning A's propensity to defect, as say the R-agent in Jurca and Faltings (2003) would be, but rather A's estimates of the future cost of a negative, at each level of influence. Thus the agents' type is different.

Note however that the above discussion assumes that the second-order reputation information is known—and available globally. Yet what motivates the provision of second-order reputation information? Why would an agent wish to share its discovery of a competitive advantage, an accurate source of reputation information, with other agents? The problem of recursion is thus still present—the provision of reputation information is motivated by the promise of access to more transactions, but there is no analogue to motivate the provision of second-order reputation.

It is not possible to adopt the same strategy as Jurca and Faltings (2003) and keep the second-order information private because the system relies on the influence of an agent being known to prospective partners in order to influence them into co-operating.

There remains one option—if there is only one reputation value contributed to by both activity in transactions and activity in providing reputation information then there is no recursion—each action results in an increase of the same score. However then we have returned to a situation in which we are assuming that agents have some underlying "quality" that determines their actions and have now come full circle. If there is an underlying quality then agents don't really make a choice depending on the partner they are currently facing.

[11] Note that this may be similar to the ontological structure of reputation discussed in Sabater and Sierra (2000), here "reputation for trustworthiness in transactions" and "reputation for accuracy generosity and trustworthiness in reputation report provision" are components of an ontologically higher concept, "reputation".

This issue requires more investigation.

3.2 Dealing with value and collusion

Another set of open research questions involves consideration of particular strategies that agents might use to benefit from a reputation system.

There is little consideration of the problems of value in the systems design literature. There is a well-known eBay case described in this online newspaper article (Kirsner, 2003) thus,

> Since most auction bidders look at a seller's feedback messages before they place a bid to determine whether the seller is trustworthy, Nelson devoted a lot of energy to creating positive feedback profiles for his various online identities. One identity of Nelson's would "sell" an item to another, and then the "buyer" would post positive feedback on the "seller." Nelson would also buy inexpensive items, like paperback books, from sellers who actually existed, hoping that they would add good feedback to his profile. He didn't care about actually receiving the books, and he regularly used a fake mailing address.
>
> Once an identity had received enough positive feedback to be considered trustworthy, Nelson would set up a "Dutch auction," in which he claimed to have a large batch of a particular item to sell. Dutch auctions allow sellers to post quantities of identical merchandise all at once, rather than item by item, and bidders can buy as many as they want. By the time buyers started complaining to Yahoo or eBay that they'd paid but never received the product, causing that particular identity of Nelson's to be suspended from selling, Nelson would have collected most of the money. In June 2000, one identity, harddrives4sale, took $32,104 from would-be buyers on Yahoo; in September, another identity raked in $12,985 on eBay.

These types of strategies have been acknowledged but not tested by the literature considered in this paper. One simple strategy would be to weight a positive or negative vote on the basis of the value of the transaction it is arising from, while another could be to allow agents to provide a prospective value with their query and have a reputation for only similar transactions returned. It is usually suggested that this situation stores too much meta-data about a transaction in the reputation database and might be considered to impact on the privacy of participants. This question appears to be a critical one for a system considering a reputation scheme—a first step would be designing a test of existing systems while using a strategy that builds a reputation on small transactions then defects on large transactions.

Collusion is an issue that requires further research. The Pagerank algorithm used by Google is said to effectively counter the ability of inter-linked groups to boost their collective reputation—they can only redistribute the reputation received from outside the group. Also of relevance is the work of Levien (2001).

3.3 Issues for distributed reputation systems

4 Conclusion and Recommendations

In my opinion the most useful introductory papers are:

- Mui et al. (2002) on reputation concepts and systems literature
- Fehr and Gächter (2002) on altruism and co-operation literature
- Miller et al. (2003) and Jackson (2000) on Mechanism Theory and its use for reputation systems. It will also be worth keeping an eye on the Harvard EECS group under David Parkes for distributed mechanism designs.
- Aberer and Despotovic (2001) as an example system most similar to the "smear both" MMAPPS proposal.
- Jurca and Faltings (2003) as an example of "layered" design which aims to ensure that the system is not compromised by its data storage or communications layer.

The issues which must be considered in any reputation system are:

- How to motivate truthful reporting without creating structures with recursive motivation problems
- How to ensure that any system that relies on behavioural insights is not exploitable by programmed agents and/or simulations—this means identifying the conditions (e.g. number of cheaters) under which a scheme would fail.
- How to ensure that the system is robust to infrastructural attacks (e.g. in data storage, presentation or communication)
- How to integrate value to protect against "bait and switch" strategies
- How to preclude the use of collusion

Note that the discussion in the literature reviewed here mainly focuses on issues of motivation and behaviour and that there is a gap in the literature on the questions of value and collusion. This gap is indicative of the limitations of simulations because simulations can only test a limited number of chosen strategies—value games and collusion have not, as yet, been a focus of such strategies.

In light of the literature the "smear both" idea is worth pursuing. The shape of the current papers provides a useful outline on how to describe and test the proposed system. While an analytical solution would be ideal if this is not possible then a simulation should be constructed around an iterated prisoner's dilemma situation with accumulated payoffs as the measure of success in the system. The study should explicitly acknowledge its definition of welfare—utilising aggregate payoffs (utility) but commenting on aspects such as distribution.

A particularly good simulation would include the ability to test situations in which an agent uses throwaway identities—this would allow experimentation with the "cost of entry" parameter. Motivation issues do not seem crucial to a "smear both" system, see discussion in Section 2.4.

The goal in such a simulation would be to detect the cheating agents and punish them by excluding them from transactions or forcing them to pay the cost of re-entry to "clean" their record.

The author intends to pursue the influence ideas more to build either a game or a simulation in which the agents predict the impact of a defection against their current opponent and use this "potential damage" in their decision to co-operate or to defect.

References

Aberer, K. and Despotovic, Z. (2001), Managing trust in a peer-2-peer information system, *in* H. Paques, L. Liu and D. Grossman, eds, 'Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM01)', ACM Press, pp. 310–317.
URL: *http://citeseer.nj.nec.com/aberer01managing.html*

> Aberer and Despotovic are focused on reputation in a distributed system and on the additional constraints that places upon incentive compatibility. They term these the "data management and the semantic level".
>
> They view reputation management as a sub-section of trust management. useful from Gnutella all the way through to when "trusting other peers is a crucial prerequisite for performing business". However, "Agents storing and processing trust related data cannot be considered as unconditionally trustworthy and their eventual malicious behaviour must be taken into account". "The reputation is essentially an assessment of the probability that an agent will cheat. Thus, the method can be interpreted as a simple method of data mining using statistical data analysis of former transactions". They use a P-Grid to store the data. A P-Grid is a distributed B-tree where each agent holds data pertaining to other agents in the system. Their earlier work describes the P-Grid and some of its desirable qualities (e.g. bootstrapablilty).
>
> Their notions of reputation doesn't currently include contextual information (and therefore doesn't include ontological dimensions) although they maintain that this would easily be possible. They talk of normalising the reputation of an agent on the global behaviour off all agents.
>
> Their most useful contribution is the separation of the problem into two questions: "the semantic question: which is the model that allows us to assess trust [of an agent based on that agents behaviour and the global behaviour standards"
>
> The second problem is,
>
> > The data management problem: how can the necessary data ... be obtained to compute trust according to the semantic trust model with reasonable effort? .... not just an ordinary distributed data management problem. each agent providing data on earlier interactions about others needs in turn also to be assessed with respect to its own trustworthiness.
>
> They go further and speak of The Global Trust model, the local algorithm to determine

trust (with the problem of unreliable witnesses and network instability) and the Data and communication management (must scale $O(logn)$).

Their substantive simulation stores complaints about agents and they study situations in which agents cheat with varying probabilities. They also utilise direct and propagated reputation.

Their proposed system makes observed behaviour available for local trust computation - i.e. the system tells them what has happened and each agent can use that in whatever way they want. Ie - the P-Grid is an accounting system of sorts. Well not really as it actually stores complaints - not information on a deal - so another cannot decide whether the complaint was reasonable or not.

But they claim that "data about the observed behaviour about all other agents and from all other agents is made directly available for evaluating trustworthiness using the local trust computation". They claim that this is different from Yu and Singh (2002) which only makes available "an abstract method".

Alessi, L. D. (1994), 'Reputation and the efficiency of legal rules', *The Cato Journal* **14**(1).

Alexander, R. D. (1987), *The biology of moral systems*, A. de Gruyter, Hawthorne, NY.

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

Azzedin, F. and Mahewswaran, M. (2002), Towards trust-aware resource management in grid computing systems, *in* 'Proc. of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'02)'.
URL: *http://www.ee.princeton.edu/r̃blee/ELE572Papers/trust_awareGRID.pdf*

Azzedin and Maheswaran propose that, rather than basing security for Grid resource sharing on sand-boxing and access control lists (and the overheads that these introduce), security and confidentiality can be increased through making the resource allocation manager 'trust aware'.

Basically they propose that to avoid a resource allocator sending confidential data to an untrusted machine within the Grid Domain the resource allocator would assess the trust level required for the task and allocate the job accordingly.

They provide a definition of trust as a "firm belief in the competence of an entity to act as expected ... " and define reputation as " an expectation of its behaviour based on other entities' observations or information about the entity's past behaviour at a given time".

They claim to be working on implementing a trust architecture but this paper is concerned only with using such trust values in a resource allocation situation.

Brainov, S. and Sandholm, T. (2002), Incentive compatible mechanism for trust revelation, *in* M. Gini, T. Ishida, C. Castelfranchi and W. L. Johnson, eds, 'Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)', ACM Press, pp. 310–311.

Clarke, I. (2003), 'Freenet next gen routing algorithmn'.
URL: *http://hawk.freenetproject.org:8080/pipermail/devl/2003-April/005008.html*

Ian Clarke discusses the FreeNet NG routing algorithm.

Essentially the plan is to have a global data structure of node response times for particular keys so that the network knows who is best to contact for a particular key. This is global results oriented routing rather than topological knowledge. There is no need for the requesting node to know how the serving node gets the key.

This is essentially a reputation for the node with the context of each key. It will be interesting to see how this global data structure is implemented and whether they are motivated to make it attack secure.

Cornelli, F., Damiani, E. and Capitani, S. D. (2002), 'Choosing reputable servents in a p2p network'.
URL: *citeseer.nj.nec.com/556137.html*

The Italian team, lead by Fabrizio Cornelli, approaches the reputation in P2P networks from the understanding that "anonymity opens the door to possible misuses and abuses by resource providers exploiting the network as a way to spread tampered with resources, including malicious programs".

They have designed an addition to the Gnutella protocol that allows peers to be rated on the basis of the quality of the files they have provided. In this sense it approaches the issue in the same manner as Schecter et al. (2003) and Kamvar et al. (2003).

They include a self-justification paragraph in which they claim to be working to "make these architectures compliant with the ethics of the user population." They point out that any reputation based extension to existing P2P protocols requires persistent identifiers (which may be opaque) - this is different from the current P2P situation and may undermine the anonymity available on the services (which is its point, in a way) unless the identifiers are truly opaque (not a hard hurdle in their and my opinion).

Servent in the title is not a typo - they claim this is the neologism for Client/Server.

Their architecture involves a step after receiving responses from other machines in the network a propagated request for reputation information about that group is made. The messages are conceived of as 'votes'. They include a 2nd order voting (called "enhanced voting") that allows reputation for voting to be assessed.

They make various efforts to confront the security implications of the protocol (i.e. false voting etc.) but do not provide either a formal proof of its strategyproofness nor a simulation of attack situations.

"We described a reputation management protocol for anonymous P2P environments that can be seen as an extension of generic services offered for the search of resources."

Crowcroft, J., Gibbens, R., Kelly, F. and Ostring, S. (2003), Modelling incentives for collaboration in mobile ad hoc networks, *in* 'Proc. of WiOpt'03'.
URL: *citeseer.nj.nec.com/586464.html*

Davis, M. S. (1971), 'That's interesting!', *Philosophy of Social Science* **1**, 309–344.
URL: *http://www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Interesting.htm*

Dragovic, B., Hand, S., Harris, T., Kotsovinos, E. and Twigg, A. (2003), 'Managing trust and

reputation in the xenoserver open platform'.
URL: *http://citeseer.nj.nec.com/kotsovinos03managing.html*

Eschenauer, L., Gligor, V. D. and Baras, J. (2003), On trust establishment in mobile ad-hoc networks, *in* 'Proc. of the Security Protocols Workshop, Cambridge, UK'.
URL: *citeseer.nj.nec.com/eschenauer02trust.html*

This paper, reached via Kamvar et al. (2003), is excellent for understanding the relationship between the security communities concept of authentication and the concepts at play in the P2P reputation community.

They conceptualise the problem as being able to assign trust to a node. On the Internet this has been accomplished by PKI, but PKI requires certain assumptions that are not applicable in a Mobile ad-hoc network (MANET). They typically assign trust via an off-line investigation of procedures (such as those from a certificate root). These are then "cached" in certificates stored locally or on certificate servers.

They develop the argument that these systems are not functional in a MANET environment which "requires protocols that are: - peer-to-peer, independent of a pre-established trust infrastructure i.e. certification authority and directory servers); - short, fast and on-line; and - flexible and support uncertain and incomplete trust evidence".

The most useful aspect of this paper is their conceptualisation of the problem in general terms, to which one can relate the problem of reputation in a P2P environment.

"We view the process of trust establishment as the application of an evaluation metric to a body of trust evidence. The outcome of the trust establishment process is a trust relation [which can then be built upon]." A policy is the local decision made on the basis of the trust establishment.

Evaluation Metric, Trust Evidence largely equate to Aberer and Despotovic (2001) understanding of the "Global Trust model and Local Algorithm" (as an Evaluation metric and a Policy) and "Data and Communication management" (as Trust Evidence). Establishing trust is the same thing as deciding whether to transact or to co-operate rather than defect (in game theoretic terms).

On MANET (and P2P) "Trust establishment has to be performed with incomplete and hence uncertain trust evidence"

The security and PKI approach of this paper brings to attention the concept of revocation - it seems that revocation has not been an issue for reputation studies because they see the evidence as being a description of a single interaction rather than the outcome of a process of investigation. However, if I receive a product which works for the first two days and then breaks, or looks right but is hollow, and vote positively then how can I revoke that voting decision?

> Evidence may be an identity, a public key, a location, an independent security assessment, or any other information required by the policy and the evaluation metric used to establish trust.

They describe basic questions regarding distribution and storage of these pieces of

evidence - it is at this stage that they turn to P2P file-sharing systems as potential storage locations for the trust evidence (which begs the question how do these networks cope with trust?). They examine Freenet and approve some characteristics but argue that since all evidence needs to be returned it would need to be adjusted. They like the locational characteristics of freenet in which commonly accessed information is stored near the requests.

They cite their intention to pursue work on swarm intelligence for trust evidence distribution, i.e.. Ant colonies behaviour (chemical trials), because "it requires exploration of the environment with reinforcement of good solutions but also regulation that allows new sources to be discovered.".

They argue that the PGP web of trust comes closest to their envisaged system in that it deals with some uncertainty but only approves chains of keys, not any evidence that could be presented.

Fehr, E. and Gächter, S. (2002), 'Altruistic punishment in humans', *Nature* **415**(6868), 137–40.

In an effort to answer the question of "how to explain human co-operation", Fehr and Gachter cite a number of theories: Theory of kin selection, Direct reciprocity in bilateral long-term interactions (Trivers, 1971), Indirect reciprocity Alexander (1987), Nowak and Sigmund (1998), Costly signalling Gintis et al. (2001) (the last two utilise reputation in their explanatory framework).

> Yet these theories do not readily explain why co-operation is frequent among genetically unrelated people, in non-repeated interactions when gains from reputation are small or absent. ... Punishment provides a solution to this problem ... Everybody in the group will be better off if free riding is deterred, but nobody has an incentive to punish the free riders. Thus, the punishment of free riders constitutes a second-order public good. The problem of second-order public goods can be solved if enough humans have a tendency for altruistic punishment, that is, if they are motivated to punish free riders even though it is costly and yields no material benefits for the punishers.

The authors designed an experiment designed to exclude all explanations for cooperation other than that of altruistic punishment. The game was similar to McCabe style investment games. A provision was made for players to punish their partners if they choose to do so. Punishment was costly - each punishment cost the punisher 1 and the punished 3.

> The act of punishment does provide a material benefit for the future interaction partners of the punished subject [in that those punished in the past co-operated more in the future] but not for the punisher. Thus, the act of punishment, although costly for the punisher, provides a benefit to other members of the population by inducing potential noncooperators to increase their investments. For this reason, the act of punishment is an altruistic act.

Finally a questionnaire with example scenarios was taken which rated anger on a likert style scale. The authors conclude that negative emotions are causing the altruistic punishment. This is because "most punishment acts [were] executed by above-average

contributors and imposed on below-average contributors". Further "punishment increased with the deviation of the free rider form the average investment of the other members" . One interesting result of the questionnaire was that free-riders indicated higher levels of expected punishment than where actually received.

The implications of this work for systems of incentive compatibility are two-fold. The first is that a reputation system is not required (and thus a persistent identifier is not required) provided that a financial punishment can be administered. Due to the 'propensity to punish' (which may be modelled as a pay-off due to 'venting anger') this element may be enough.

If considered in the context of reputation systems it could imply that only a negative counting system is required (where negatives have some credible cost to the receiver) - and that there will not be a public good issue/ free-riding problem in the motivation of the negative vote provision.

There are, however, some problems in the relevance of this work to the MMAPPS and other agent-based problems. One confounding element is whether this altruism sticks if it becomes known that the punishers do worse in the long-run I consider it likely that when the costs of the punishing strategy are known the 'propensity to punish' will decline.

Can a system based in the notion of a motivation driven by emotion be utilised in an agent context? This theory boils down to 'people get angry and are prepared to pay a cost to punish'. Can this be reduced to a utility parameter for satisfying anger? If so there are two sources of utility in this game, money and "satisfaction of anger". They are only exchangeable when anger is present ...

The danger from an agent systems builder's point of view is that we cannot rely on the presence of actual humans in the system - if an agent is acting on a human's behalf it is unknown whether a person would pre-program an agent to act on 'anger' - given its status as a costly strategy. Furthermore if an agent is not acting directly for a human would that agent (working through simulation) choose not to punish...

Friedman, E. and Parkes, D. (2003), Pricing WiFi at Starbucks—issues in online mechanism design, *in* 'Proc. Fourth ACM Conf. on Elec. Commerce (EC'03)'.
URL: *http://www.eecs.harvard.edu/econcs/pubs/online.pdf*

Fudenburg, D. and Tirole, J. (1991), *Game Theory*, MIT Press, Cambridge, MA.

Gintis, H., Smith, E. A. and Bowles, S. (2001), 'Costly signalling and cooperation', *Journal of Theoretical Biology* **213**, 103–119.

Gintis et al are writing in the vein of literature focused on explaining co-operative human behaviour within a biology and social biology context. In this sense they are similar to Fehr and Gächter (2002) and Nowak and Sigmund (1998). Their examples are often taken from hunter gatherer studies and animal behaviour studies. This paper is a formal game theory formation of the notion of costly signalling that had been developed only descriptively before this paper.

The notion of evolutionary fitness in these studies is a critical judgement of success -

one gathers resources to survive and procreate - this success metric is equivalent to the accrual of payoffs in the market simulations such as Jurca and Faltings (2003).

"Some hunters consistently provide more than others while sharing equally in the catch" - These altruistic players are acting in this fashion to effective signal their skills in hunting, endurance, generosity etc. to build reputation. However this may be better expressed as status in that it is explicitly signalling about attributes that include out of context attributes - their term is "high quality". Defined as, "genetic or phenotypic attributes that are difficult for others to assess directly, yet have important effects on the payoffs from social interactions with the signaller".

This is a fairly subtle notion that relies on the assessment of meaning in actions - we see generosity in one context and impute that that means generosity (and possible a host of other correlated features) is likely in another context. One who is generous is likely honest etc. I think that this is closer to the 'smudged' role that reputation plays in the real world. Compare this to the stricter notion of reputation in different contexts outlined in Sabater and Sierra (2000) (which is at least hierarchical) and in Aberer and Despotovic (2001); Yu and Singh (2002).

Their model is also applicable to public good provision, such as defence or punishment of free-riding. These altruistic actions create other opportunities for the provider - thus resulting in more evolutionary fitness.

The benefit derived through this signalling is that people are more likely to engage in other interactions with that person that are beneficial to both parties. In the context of an online market this might mean that the provision of reputation reports (i.e. good reputation for providing reputation reports - 2nd order) would lead to more opportunities to transact.

The 'leaky' role of reputation here potentially supports the suggestion that co-operation in trades could be the source that builds trust in the 2nd order recommendations of that agent. That is - by combining reputation contexts between 1st and 2nd order we can avoid the re-cursiveness problem (1st order rep motivates market co-operation and 2nd order rep motivates provision of 1st order rep reports, but what motivates the provision of 2nd order rep reports? and so on).

Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology* (91), 481–510.

Grimes, K. (2003), 'To trust is human', *New Scientist* **10 May**.

An article summarising the developing field of NeuroEconomics that is arising to address the questions raised by experimental economics. The school is associated with Nobel prize winner Vernon Smith and Paul Zak.

The emergence of co-operative behaviour in iterated prisoners dilemma games have lead some to posit the existence of trust. This is typically understood as the existence of a social norm but neuroeconomics posits that there is a physiological imperative to trust - trust behaviour releases Oxytocin which is a feel-good hormone. It is hypothesised that this hormone is part of the reason for co-operative behaviour which is hard to explain with the 'all for myself' model of evolution.

The research involved measure levels of hormones released during a McCabe style investment game which is meant to involve two way trust - from the investor and the bank.

> Trust in our species therefore appears to be driven by an emotional 'sense' of what to do, rather than a conscious determination. (Zak).
>
> Zak's interpretation of his findings poses a challenge to economic tenets like the Nash equilibrium that assume we consciously and rationally seek to maximise personal profits. These models see human motivation as a kind of "lucid greed", transparent to the introspection of oneself and others. Observed cooperation is then explained as an emergent property of culture and society, imposed from above on the natural selfishness that is the human default motivation. Zak's work suggests, in contrast, that social cooperation can arise as a primitive impulse in ancient brain areas - an impulse that successfully contests the lucid greed generated by more recently evolved brain regions.

Smith and McCabe believe that cooperation depends strongly on the opinion about the moves of others. "Smith and McCabe conclude that the decision to trust depends on projecting one's own co-operative intentions onto another person."

Jackson, M. (2000), 'Mechanism theory'.
URL: *citeseer.nj.nec.com/jackson00mechanism.html*

Jurca, R. and Faltings, B. (2003), An incentive compatible reputation mechanism, *in* 'Proceedings of the IEEE Conference on E-Commerce', Newport Beach, CA, USA.
URL: *http://liawww.epfl.ch/Publications/Archive/Jurca2003a.ps.gz*

> This paper considers distributed reputation systems from the perspective of incentive compatibility.
>
> They conceptualise the problem as assisting with an iterated prisoners dilemma in which agents wish to play a game with a partner that will co-operate and avoid playing the game with a partner that will defect. They are coming from the agents paradigm focused on the development of e-commerce.
>
> Their principle contribution is to point out that the motivation for an agent to provide reputation information itself is not obvious and that there may in fact be motivations for an agent to either not report reputation information, or to provide false reputation information. Not reporting or providing false information may be of advantage to an agent in a competitive situation - they do not consider the existence of a 'norm of punishment and complement' that appears to be borne out in the Ebay literature.
>
> Furthermore the model includes a currency for the reputation market which is not exchangeable for real currency yet there is no consideration of the nature of such a currency nor how transactions would occur through it.
>
> Their solution creates a secondary market for reputation information through the creation of 'R-agents' who buy and sell reputation information. The market is structured in a fashion such that it is more profitable to provide truthful reputation reports. This structure is sketched below. However it is clear that there is a secondary game occurring between the agents and the R-agents - how is an agent to trust the R-agent?

They propose to bootstrap this trust by having agents store their local experiences with the (expected to be low) numbers of R-agents. It is not clear what the R-agents motivations are vis-a-vis truthful reporting of their knowledge.

The reputation information market is structured towards incentive compatibility:

> We propose a simple payment scheme that makes it rational for agents to truthfully share the reputation information. The basic idea is that: R-agents pay the reputation report of agent A about agent B only if it matches the next report submitted about B.

Lying agents, being paid less for their false reports, eventually have no "reputation currency" left and thus cannot participate in the reputation market, being unable to buy reputation reports and, in the design of the system, thus unable to submit more false reports. It is not clear how reputation currency is initially gained.

It is clear that "incentive compatibility" is differentiated from strategyproofness: "Clearly, an increasing number of lying agents will affect the incentive-compatibility property of the mechanism, as it can no longer be guaranteed that truthful reports are paid with higher probability" They then derive a theoretical threshold for the percentage of lying agents that will break the system. Their answer is that if there are greater than 50agents lying then the system will loose its incentive compatibility.

The paper suffers, in my opinion, from the temptation to solve the problem through indirection. Yet how can one trust transactions in this indirect market? Their answer is that one uses personal experiences with the R-agents - thus trust in the broader market is bootstrapped through a higher-trust, lower-number of players market. The R-agents are conceived to be under some type of centralised regulation in that they are to be non-profit and simply cover their costs. The nature of this regulation is not clear.

Any attempt to use indirection to solve the problem simply shifts the problem - yet trust exists in the real-world - how can this real-world trust be leveraged. PGP web of trust uses government ID (which is far from strategy-proof) to bootstrap that system.

They claim to be implementing this system on the agentcities agent framework - which is 5th framework funded.

Kamvar, S. D., Schlosser, M. T. and Garcia-Molina, H. (2003), The eigentrust algorithm for reputation management in P2P networks, *in* 'Proceedings of the Twelfth International World Wide Web Conference (WWW)'.
URL: *citeseer.nj.nec.com/article/kamvar03eigentrust.html*

Kamar, Schlosser and Garcia-Molina, from Stanford, present a reputation system focused on preserving the quality of files in a file sharing network. Their notion of reputation is a binary indication of whether a file is authentic or inauthentic. The attack situation their are addressing is a situation similar to the RIAA or Madonna attempting to 'poison' the P2P networks by uploading corrupted files.

Their system uses these quality votes to assign reputation to peers (using an opaque identifier - here they suggest Gnutella username). The reputation scores are then used to probabilistically determine which peers, from a set of responses to a query, a peer

will download a file from. They adjust this probability to avoid the network converging on a single most trusted peer.

The system suggests the utilisation of a Distributed Hash Table for data storage and presents an algorithm for distributing the calculation of global trust scores for each peer. The trust, therefore, is a single global aggregation of the experiences of all of the peers in a network. They state that in a future paper they will consider a situation in which trust scores are used to exclude untrusted peers from the trust calculations.

They do not attempt strategyproofness but consider a number of attacks which operate via the data storage layer and through manipulation of the semantic votes themselves. Their simulation, modelled on real world characteristics of the Gnutella network, compares a network without a trust system with their system under four situations of attack by malicious peers and groups of malicious peers. They do not simulate attacks on the Data storage layer, having attempted (but not proved) to dispel these in the design of the DHT.

Useful quotes:

Reputation is useful to ensure that "peers obtain reliable information on the quality of resources they are receiving" Cornelli et al. (2002)

Reputation is useful for resolving the "complete lack of accountability for the content a peer puts on the network"

They claim that global reputation values are useful for "Isolating malicious peers" and "Incenting Freeriders to Share" - The second is to be implemented through greater bandwidth - but it is not clear how that might work ...

Kirsner, S. (2003), 'Catch me if you can', *Fast Company* .
URL: *http://www.fastcompany.com/magazine/73/kirsner.html*

Levien, R. (2001), Attack-resistant trust metrics, PhD thesis, UC Berkeley.
URL: *http://www.levien.com/thesis/*

Marinoff, L. (1992), 'Marinoff, l. (1992) maximizing expected utilities in the prisoner's dilemma', *The Journal of Conflict Resolution* **36**(1), 183–216.

Michael Schillo, P. F. and Rovatos, M. (2000), 'Using trust for detecting deceitful agents in artificial societies', *Applied Artificial Intelligence* **Special Issue on Trust, Deception and Fraud in Agent Societies**.
URL: *citeseer.nj.nec.com/463797.htm*

Milgrom, P. R. and Roberts, J. (1982), 'Predation, reputation and entry deterence', *Journal of Economic Theory* .

Miller, N., Resnick, P. and Zeckhauser, R. (2003), Eliciting honest feedback in electronic markets, *in* 'Working Paper for the SITE02 workshop (updated)'.
URL: *http://www.si.umich.edu/~presnick/papers/elicit/*

"This paper proposeds a payment-based system to induce honest reporting of feedback" using "proper scoring rules".

The paper begins with fairly anecdoctal observations regarding the lack of motivational incentives for individuals to provide truthful feedback information. More strongly they draw on their earlier empirical work on eBay feedback (Resnick and Zeckhauser, 2003).

They argue that "... a mechanism that will elecit hones feedback from agents ... is an instance of a more abstract problem, that of indicing agents to reveal private information." This approach places them well within the mechanism design approach discussed in Jackson (2000).

Their approach involves a centralized mechanism. This is regular in Mechanism Theory and is the element that the work of Parkes (2001) is attempting to distribute and make computationally efficient. The system "is based on typing the payments to buyers to the informativeness of their evaluations ... the payment for an evaluation needs to reflect the degree to which the evaulation agrees with the evaluations of others." In this sense it is similar to Jurca and Faltings (2003) and other schemes. The crucial question of judging similarity is handled by the centre - which has global information.

The sellers have fixed quality (although they claim in a footnote that "Future work should allow sellers to respond to their developing reputations"). This raises some significant questions as to the usefulness of the system - it can only act as a signal, not a sanction.

The paper cites Resnick et al. (2000) three characteristics of a reputation system:

1. help people decide who to trust
2. encourage trustworthy behaviour and appropriate effort
3. deter participation by those who are unskilled or dishonest

When sellers have no ability to alter their behaviour reputation acts as a signal only - thus the system described in this paper will not achieve point 2 - it cannot encourage trustworthy behaviour - this is determined exogenously. Their system makes truthful reporting a Nash equilibrium and there are balanced transfers (i.e. the budget is balanced - the same amount is paid to the centre as is paid out). They use an indirect system that recalls a 2nd price auction in which, in their example, player 3 pays player 1 on the basis of player 2's score. "Thus for each player a second player's signal is used to determine incentive payments, and a third player makes the payments." There is a long formal analysis rooted firmly in Mechanism Theory.

Moreton, T. and Twigg, A. (2003), Trading in trust, tokens and stamps, *in* '1st Workshop on the Economics of P2P systems', Cambridge University.
URL: *http://www.sims.berkeley.edu/research/conferences/p2pecon/papers/s2-moreton.pdf*

Mui, L., Halberstadt, A. and Mohtashemi, M. (2001), Ratings in distributed systems: A bayesian approach, *in* '11th Workshop on Information Technologies and Systems (WITS)'.
URL: *citeseer.nj.nec.com/521476.html*

Mui, L., Mohtashemi, M. and Halberstadt, A. (2002), Notions of reputation in multi-agents systems: a review, *in* 'Proceedings of the first international joint conference on Autonomous agents and multiagent systems', ACM Press, pp. 280–287.

This contains a useful taxonomy of types of reputation - unfortunately the diagrams don't come out quite right.

They survey reputation in Economics, Academic publishing, Computer Science, Evolutionary Biology, Sociology and Politics.

The Economists they survey include Fudenburg and Tirole (1991) which is a classic work on game theory that "Economists often interpret the sustenance of co-operation between two players as evidence of "reputation effects" ". They therefore take from economics the approach of the Prisoner's Dilemma and ascribe the emergence of co-operation to the development of reputation. This has also been called the development of trust.

The second group of Economic literature that they discuss is the role of firm's reputations in restricting entry to markets Milgrom and Roberts (1982). Finally they discuss Tadelis (1999) discussion of reputation as a tradable asset.

From computer science they discuss (Zacharia and Maes, 1999) and the Histos model in which reputation scores are different depending on the context in which the query is made. Mui's earlier work Mui et al. (2001) which is discussed later in the modelling part of the paper uses a Bayesian statistical analysis to infer a reputation for an agent based on propagated ratings and the reputation of the neighbours.

In Sociology they refer to the social network analysis and the concept of prestige and Granovetter (1985)

The Typology they develop is Group and Individual Reputation. Individual Reputation is broken into Direct and Indirect reputation (based on word of mouth).

Direct Reputation is (somewhat dubiously) broken into Observed and Encounter Based reputation. Observed reputation is a little unclear in how it is distinguished from Indirect (reported) reputation.

Indirect Reputation is broken into Prior-derived reputation (default values), Group-derived reputation (inferring based on group membership) and Propagated reputation. Derived Reputation is considered by Mui et al. (2001), Sabater and Sierra (2000) in his REGRET system and Michael Schillo and Rovatos (2000).

They introduce an interesting motivational discussion from evolutionary game theory. Trivers (1971) as having developed the concept of reciprocal altruism and discuss Alexander (1987) extension of this to a concept of indirect reciprocity. They use these concepts of reciprocity in their game.

They set up a prisoner's dilemma simulation in which reciprocating agents are matched against always-defecting agents. The reciprocating agents are Tit-For-Tat in that they will do whatever the other did last round. In the first round they have both regular TFT (which co-operate in the first round) and Reputation TFTs which base their first round decision on the reputation of the partner.

The partner's reputation can be either encounter-driven individual reputation, Observed individual reputation (in which the agent watches 10 others and uses an average of their behaviour to estimate the probability of its partner defecting), Group-derived reputation (each agent belongs to a group and the group reputation is used to decide the first move) and propagated reputation using the Mui et al. (2001) reputation propagation technique.

An agent is selected at random and another to play against - this occurs a number of times which defines a generation. At the end of the generation agents beget similar off-spring in proportion to their fitness (i.e. their success as measured by the payoffs in the Prisoners Dilemma).

Their results show that there is a significant advantage when group, propagated (d=1) and propagated (d=3) are used. The metric used is the number of Encounters Per Generation. The baseline is a game with only Always-Defecting and TFT agents - when a high number of EPG is reached the TFT strategy begins to dominate because agents are meeting each-other more.

The TFT was replaced in turn by each of the different RTFT agents. The encounter-derived agents took 14000 EPG to dominate while the propagated (d=3) agents took less than 1000 to dominate over the always defect agents.

Nowak, M. and Sigmund, K. (1998), 'The dynamics of indirect reciprocity', *Journal of Theoretical Biology* **194**, 561–754.
URL: *http://www.ptb.ias.edu/nowak/pdf/JTB98b.pdf*

Nowak and Sigmund develop a theory of indirect reciprocity which is the closest analogy to the underlying assumptions behind the push for reputation systems.

Their basic proposition is that "Discriminators provide help [co-operate] to those individuals who have provided help. Even if the help is never returned by the beneficiary, or by individuals who in turn have been helped by the beneficiary, discriminating altruism [can be a successful evolutionary stable strategy]"

They place their work in the line of Trivers (1971), which they term a "seminal paper", and that of Alexander (1987) crediting Alexander with coining the phrases "indirect reciprocity". They quote Alexander (1987) indirect reciprocity "involves reputation and status and results in everyone in the group continually being assessed and reassessed". In the final section they also reference Pollack and Dugatkin (1992) who suggested a strategy of "Observer Tit for Tat" in which the first move was based on the action of the co-player in the previous game (this is similar to the Reputation TFT in Mui et al. (2002)).

> In this case, one does not expect a return from the recipient (as with direct reciprocity), but from someone else ... a donor provides help if the recipient is likely to help others' (which is usually decided on the basis of experience, i.e. according to whether the potential recipient has helped others in the past).

Their analytical game is structured as follows. Each player has an image score (reputation) which is incremented by 1 if he helps a player requiring help, and decremented by 1 if he refuses to help. The Discriminator strategy sets a value, k, which is an image threshold for assisting players in need. This strategy is matched against allways-help and never-help strategies. By helping, therefore, "they increase their score and are, therefore, more likely to receive help in the future".

They express one finding as, "We show that the probability q that a player knows the score of another player must exceed $\frac{c}{b}$" where c is the cost of assistance while b is the

benefit to the assisted, with the assumption that $c < b$, and therefore the assistance is socially valuable.

Note that the game is such that the activity is observed by other discriminators - who then bases their actions on their knowledge of the image score of the player if they encounter them. There is no propagation of this information, and therefore there is no need to motivate presentation of image information. It also assumes that players "know what they see" such that a member of the group is easily able to assess whether a player did, in fact, help a player in need and does not have to rely on the reports of the helper or the helped. Finally it does not consider throwaway identities - the evolutionary game framework in general does not do this - survival is not a game that can be opted out of, nor is one's genotype exchangeable.

They do move to consider a situation of imperfect information at the end of their paper, "Even in small groups, where everyone knows everyone else, it is unlikely that all group members witness all interactions. Therefore each player has a specific perception of the image score of the other players. They argue that the outcome of that analysis "looks exactly like Hamilton's rule for altruism through kin selection, except that the coefficient of relatedness, k is replaced by the probability to know the co-player's score, q".

Finally they discuss Sugden (1986) which inverts the strategy. A player is born with good standing, which they keep as long as they help other players in good standing. defection is then possible against players not in good standing.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1998), The pagerank citation ranking: Bringing order to the web, Technical report, Stanford Digital Library Technologies Project.
URL: *citeseer.nj.nec.com/page98pagerank.html*

Parkes, D. C. (2001), Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency, PhD thesis, University of Pennsylvania.

Pettit, P. (1986), 'Free riding and foul dealing', *The Journal of Philosophy* **83**(7), 361–379.

An investigation of many-player co-operation games modelled as N-party prisoner's dilemmas.

> The free-rider problem is the predicament ... of productive cooperation. It arises when everyone is better of if each contributes to a certain cause - by effort, finance, restraint, or whatever - than if no one does so, but when no one's contribution is likely to make a difference sufficient to repay him for the cost involved. *p 361*
>
> The foul-dealer problem ... is the predicament of peaceful coexistence. This is the problem of how to persuade people individually to disarm themselves of some instrument of offence when, albeit, they are better off under universal disarmament, still they are each exposed by self-disarmament to the worst prospect of all - that of being a defenceless victim of another's aggression; and, equally, if they refuse to disarm they are each eligible for the best possibility of all - that of being an unopposed aggressor. *p 362*

Formally:

> If a many-party dilemma is one in which no co-operator is made worse off by a lone defector than he would be under universal defection, then we have a type A dilemma ... If it is one in which the lone defector plunges some co-operator or co-operators below that base line [of universal defection] then it is a type B dilemma

p 365

"Thus the lone free rider does not plunge anyone below the baseline of universal defection. The free-rider problem is a type A prisoner's dilemma" p 370

> [For the foul-dealer the] benefit .... must be enjoyed, not just by the efforts of others, as in the free-rider case; it must be enjoyed directly at some other's expense. .... there is more to be enjoyed [by the defector] than just the share of the good procured by the efforts of others. .... [The co-operators] must present the lone defector with the opportunity to exploit one or more of the co-operators, letting him enjoy some predatory advantage *p 372*

> The advantage may be immediate, as in queue jumping or price cutting, or it may come from repute in the eyes of others. This repute will involve the attainment of a higher standing on some comparative metric, whether a metric with absolute significance, like beauty or strength or virtue, or one of an intrinsically relative kind such as status. *p 373*

It is interesting to consider whether the 'context slippage' property of reputation means that it has some 'absolute significance' - does this explain the power of the 'do not take my name' from The Crucible. By 'context slippage' I mean that reputation from one context appears to be used in other contexts, rightly or wrongly.

The collective good is two dimensional, "not now the aspect of a good to be shared with others; rather that of an advantage that other's cede".

"The free rider seeks to benefit by the efforts of others, the foul dealer to benefit at their expense" p 374

He goes on further to argue that these situations have a strategic difference.

Pollack, G. B. and Dugatkin, L. A. (1992), 'Reciprocity and the evolution of reputation', *The Journal of Theorectical Biology* .

Resnick, P. and Dellarocas, C., eds (2003), *Online Reputation Mechanisms: A Roadmap for Future Research.*
URL: *http://www.si.umich.edu/ presnick/reputation/symposium/ReportDraft1.doc*

Resnick, P. and Zeckhauser, R. (2003), *Advances in Applied Microeconomics*, Vol. 11, Elsevier Science, chapter Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce.
URL: *http://www.si.umich.edu/~presnick/papers/ebayNBER/*

Resnick, P., Zeckhauser, R., Friedman, E. and Kuwabara, K. (2000), 'Reputation systems: Facilitating trust in internet interactions', *Communications of the ACM* **43**(12), 45–48.

Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K. (2002), The value of reputation on eBay: A controlled experiment, Technical report, School of Information, University of Michigan.
URL: *http://www.si.umich.edu/ presnick/papers/postcards/*

Sabater, J. and Sierra, C. (2000), 'Regret: A reputation model for gregarious societies'.
URL: *citeseer.nj.nec.com/sabater00regret.html*

Sabter and Sierra have developed a model that contributes three useful concepts.

Reputation models only make sense in the context of a negotiation - they are an assistant in the process of deciding whether to trust a partner.

Reputation has three dimensions. They call these the Individual, the social and the ontological. The individual is the personalised - it is the sum of an individuals experiences with an opponent. The social dimension is the sum of the experiences with an opponent of members of that group.

The ontological dimension of reputation is that reputation is a composite of different types of reputation - e.g.. The reputation as a tour operator is a combination of hotel, flights, food, and possibly weather. Individuals have an ontological structure which reflects the weights that individuals place on these composite aspects of reputation.

They develop these observations into a model and run simulations for both an ontologically simple situation and an ontologically complex situation. The simple situation models a situation in which a high-reputation seller changes strategy to act as a low reputation seller (i.e. begins to defect) - their system identifies this behaviour faster than the current Amazon ratings or the Histos system. This is largely due to the "forgetting" that the system does (i.e. it only considers the last 100 transactions).

They are not concerned with extra requirements of distributed systems such as Data storage, nor are they, in this paper, focused on strategy-proofing such a system.

Schecter, S. E., Greenstadt, R. A. and Smith, M. D. (2003), Trusted computing, peer-to-peer distribution, and the economics of pirated entertainment, *in* 'The Second Workshop on Economics and Information Security'.
URL: *http://www.eecs.harvard.edu/s̃tuart/papers/eis03.pdf*

Schecter, Greenstadt and Smith, PhD students at Harvard EECS, walk through the implications of Trusted Computing for the entertainment industries efforts to combat unauthorised copying. Their argument makes trusted computing seem to be a response to unauthorised copying that is, at best, unlikely to work and at worst, problematic for the industry's goals.

They argue that there are two costs involved in the distribution of unauthorised copies - extraction (e) and distribution (d). In the digital era e has been reduced to near zero by cheap consumer extraction (CD ripping) while d is being reduced by P2P file-distribution systems.

The industry response to the reduction of e has been trusted computing. Trusted hardware running and trusted OS running trusted software will have its ripping capabilities eliminated. This will be guaranteed to the media companies by the hardware and software producers. In theory then, the industry will encrypt files such that they could

only be played on this trusted equipment and be assured that they cannot be copied to untrusted formats.

They point out that despite this infrastructure since the content must, at some point, be viewed or listened to by humans. If it can be communicated to humans it can be captured in this analogue phase. However the trusted computing infrastructure may raise e.

In the abstract the trusted computing infrastructure allows remote groups to be sure that the system they are speaking with is acting in a known way. The contribution of this paper is to point out that this cuts both ways - P2P networks can, through the creation of alternative chains of trust, be sure that the system the client is communicating with is identical to itself.

This is important due to the strategies pursued against P2P networks (and aiming at increasing the cost of d). These strategies are: 1. Attacking confidentiality and suing through the legal system, 2. attacking integrity (poisoning the system with low quality items) and 3. Attacking the availability of the network by understanding the network protocols and utilising altered client software to take advantage of its structure.

This is where they discuss reputation - reputation is understood as a method to defend against the

If a P2P network can guarantee, through its own chain of trusted computing, that all clients are running the same software then defences against all three attacks are relatively simple to envision. Thus Trusted Computing may come back to haunt the entertainment industry.

Of course this depends on the ability of users to specify their own trusted roots for software - if the OS or hardware vendors restrict this ability then such a scenario would not obtain. Yet it seems very likely that consumer fear of a Vendor controlled computing environment will create market pressures that will push against this ability.

An interesting round of the political economy of digital reproduction.

Actually this is a classic "interesting paper" Davis (1971) in that it refutes (inverts) what the community thought it knew about Trusted Computing. It is a combination of a type vi(b) interesting paper that comments on Function - "What seems to be a phenomenon that functions effectively as a means for the attainment of an end is in reality a phenomenon that functions ineffectively." and a type vii(a) "What seems to be a bad phenomenon is in reality a good phenomenon."

Shirky, C. (2003), 'A group is its own worst enemy'.
URL: *http://www.shirky.com/writings/group_enemy.html*

Groups grow and require structure - it will come. Attacks on core values from within.

Three things to accept:

1. You cannot separate social and technical issues.
2. Members are different from users - a hierarchy based around a core of regular users will emerge.

3. The core group, by dint of caring, get more rights than others - these can trump atomistic rights in some circumstances.

Four things to design for:

1. Handles the user can invest in. Reputation systems are in the brain, "Almost all the work being done on reputation systems today is either trivial or useless or both, because reputations aren't linearizable, and they're not portable."

   > And when the community understands that you've been doing it and you're faking, that is seen as a huge and violent transgression. And they will expend an astonishing amount of energy to find you and punish you. So identity is much less slippery than the early literature would lead us to believe.

2. Way for their to be members in good standing - Way for good works to be recognised.
3. Barriers to participation - 'segmentation of capabilities' ((could this be opportunities to interact?))
4. Ways to spare the group from scale. "The value is inverse to the size of the group". When MetaFilter gets publicity they stop taking new members. (Charging for the WELL does all of 3 and 4).

Sugden, R. (1986), *The Evolution of Rights, Co-operation and Welfare*, Blackwell, Oxford.

Tadelis, S. (1999), 'What's in a name? reputation as a tradeable asset', *American Economic Review* **89**(3), 548–563.

Trivers, R. L. (1971), 'The evolution of reciprocal altruism', *Quarterly Journal of Biology* (46), 35–57.

Yu, B. and Singh, M. P. (2002), 'Distributed reputation management for electronic commerce', *Computational Intelligence* **18**(4), 535–549.

Zacharia, G. and Maes, P. (1999), Collaborative reputation mechanisms in electronic marketplaces, *in* 'Proceedings of the 32nd Hawaii International Conference on System Sciences'.

Zak, P. J., Kurzban, R. and Matzner, W. (2003), The neurobiology of trust, *in* 'Proceedings of the 2003 Economic Science Association Conference'.
URL: *http://www.peel.pitt.edu/esa2003/papers/zak_neurobiologytrust.pdf*