

Modelling the Role of Trust in Social Relationships

ALISTAIR G. SUTCLIFFE, University of Manchester

DI WANG, Khalifa University

ROBIN I.M. DUNBAR, University of Oxford

A social trust model is presented for investigating the social relationships and social networks in the real world and in social media. The results demonstrate that multi-level social structures, with a few strong relationships, more medium ties and large numbers of weak ties emerge in an evolutionary simulation when wellbeing and alliances are rewarded with high levels of social interaction. 'Favour-the-few' trust strategies were more competitive than others under a wide range of fitness conditions, suggesting that the development of complex, multi-level social structures depends on capacity for high investment in social time and preferential social interaction strategies.

Categories and Subject Descriptors: Human-centered computing; Collaborative and social computing; Collaborative and social computing systems and tools

General Terms: Simulation tools

Additional Key Words and Phrases: Trust, social relationships, social media, social brain hypothesis

1. INTRODUCTION

Trust underpins human social relationships, binding people in the anticipation of benevolent, reciprocal interaction, which forms the basis for friendship [Brown and Brown 2009]. As computer-aided social networking has become an established Internet technology, some have questioned whether social network sites might change the nature of human social relationships [Ellison et al. 2007].

Trust in real life is governed by social exchanges in which the honesty (or deceit) of the counter-party can be assessed through the 'rich media' of social presence, verbal and non-verbal communication [Mitchell 1997]. However, social media may change the nature of inter-personal exchange in two ways: first, the Internet affords communication that reduces the barriers of time (i.e. asynchronous, persistent media) and space; secondly, social information is broadcast, at least partially, so it can be easily monitored, for example, via personal profiles, wall posts and status updates in Facebook and other social networking sites [Sutcliffe et al. 2011]. Reducing cost and increasing the scope of communication potentially allows people to change how they communicate and manage their social relationships. For example, they could choose to split their social time among many more people, concentrating on weak ties which might provide them with more information and resources

The preparation of this paper was supported by ESRC/EPSRC Cognitive Foresight Programme project TESS (Developing Theory for Evolving Socio-technical Systems). RD's contribution was also supported by the British Academy Centenary Project Lucy to Language and the EU-FP7 SocialNets and ICTe-collective projects.

Authors' addresses: A.G. Sutcliffe, Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK; D. Wang, EBTIC Khalifa University, Abu Dhabi; R.I.M. Dunbar, Department of Experimental Psychology, University of Oxford, UK.

Permission to make digital or hardcopies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credits permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© Alistair G. Sutcliffe, Di Wang and Robin I.M. Dunbar 20xx Modelling the Role of Trust in Social Relationships, *ACM*

DOI:

[Granovetter 1973]. Alternatively, they may concentrate on fewer people in more intense relationships, as observed in empirical studies on friendship [Hays 1989; Wellman et al. 2006]. Since trust is probably a function of the frequency of social interaction in relationships, Internet technology could mediate fundamental changes in human social behaviour. Although the empirical studies to date suggest that this is not so, and that social networking sites function to supplement rather than supplant real-world social relationships [Lampe et al 2006; Ellison et al. 2007; Joinson 2008; Roberts and Dunbar 2012], it is still too early to assess long-term effects. Computational modelling of human social relationships may therefore provide insights into possible impacts of social networking via the Internet.

Computational models of social behaviour have focused on cooperation and the evolution of altruism [Janssen 2006]; and on swarm intelligence and coordination in social behaviour [Bonabeau et al. 1999]. Agent-based social network models have implemented limited causal theories for social behaviour, although these are usually encoded as rules, reasoning over agents' states and attributes, such as reputation or trust values [Gilbert and Troitzsch 2005]. In this paper we explore how the causal mechanism of trust can lead to the formation and maintenance of social relationships, since trust has been acknowledged as a major influence on the strength of friendship [Hays 1985; Ostrom 2002; Oswald et al. 2004].

We explore the causal processes underlying Dunbar's Social Brain Hypothesis (SBH) [Dunbar 1998, 2010], an evolutionary social psychological theory that proposes that humans form relationships with different levels of intensity, from very best friends to close and then casual friends. It is a suitable focus for computational modelling of the effect of Internet-mediated social behaviour, which may well afford people new choices in how they spread their social time among more, less intimate or fewer, intimate relationships. SBH synthesises evidence from the fossil record with studies of current human social behaviour within a primate-wide cognitive context [Dunbar and Stiller 2007; Dunbar 2010]. Based on a general relationship for social group size and brain volume in primates, SBH predicts a typical group (or community) size for humans of 150 individuals [Dunbar 1998].

Different intimacy levels emerge from the fact that social time is inevitably limited [Lu et al. 2009]; hence individuals may focus their social interactions on a few more intense relationships at the expense of many, more diffuse ones. The innermost two layers have been identified as the support clique [Dunbar and Spoors 1995; Dunbar and Stiller 2007] and the sympathy group [Buys and Larson 2008; Dunbar and Stiller 2007], followed by an affinity group of 50 and an active network of 150 individuals. Although SBH predicts an intimacy-based structure in social networks, it does not, however, give any current-day, cognitive account of why social structures emerge from human behaviour.

If the tenets of SBH are true, it must be acting through a cognitive mechanism that might generate such a distribution of relationships within an individual's lifetime. We argue that trust provides a suitable candidate mechanism, since trust is a cornerstone in friendships and the formation of relationships. The hypothesis of social layers of increasing intimacy is also supported by the literature in the psychology of friendship [Hays 1989; Wellman et al. 2006]; however, no mechanism describing how such layers might arise has been proposed. Trust has been implicated in development of cooperation and social relationships [Ostrom 2002], and this motivated the social trust simulation reported in this paper, which builds on the model of trust we developed for modelling social relationships [Sutcliffe and Wang 2012]. In that paper we reported a preliminary simulation using the trust model to investigate the structure of relationships in ego-centric social networks proposed by SBH.

In this paper we report further simulations using the trust model which investigate how a genetic basis of SBH might evolve given different strategies for social interaction. The paper is structured as follows: after a discussion of related work, a conceptual model describing the role of interaction, trust formation and the development of social relationships is proposed. The conceptual model is elaborated as a computational model in section 4, which is then used in the experiments described in section 5.1 investigating the effect of interaction parameters on trust formation, and in section 5.2 on the effect of fitness criteria on the development of social network structures. The first set of experiments test the assumptions underlying the model and investigate the conditions under which trust might create relationships at different levels of intensity. The results of simulations varying fitness conditions are then reported in section 5.2. The paper concludes with a discussion of the computational trust model as a mechanism for studying social relationship formation and regulation.

2. BACKGROUND: TRUST IN SOCIAL RELATIONSHIPS

2.1 Psychology of Trust and Friendship

In psychology, trust in the context of interpersonal relationships is defined by Rotter [1971, p. 444] as “an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied on”. However, trusting an individual places the trustor at risk [Kramer, 1999], and therefore relies upon encountering consistent and benevolent behaviour in others [Larzelere and Huston 1980]. Cannon, Doney and Mullen [1998] defined trust as a willingness to rely on another party, and therefore to make oneself vulnerable to that party. McCabe [2002] argued that trust evolved to enhance collaboration through improved ability to assess trustworthiness and detect deceit in social interactions. As more evidence accrues through repeated interaction, trust may change from a calculative mode in which an individual’s trustworthiness is assessed on limited evidence to relational trust based on social experience [Ostrom 2002].

Social relationships among non-kin involve reciprocation of commitment or exchange relationships that necessarily rely on trust: trust that other members of one’s network will cooperate or reciprocate when required to do so, and trust that none of these individuals will deliberately undermine or abuse that relationship [Dunbar 1998]. Theories and studies of human friendship [Brown and Brown 2006; Hays 1985; Oswald et al. 2004] note the importance of reciprocity and the exchange of benevolent acts in building social relationships and, implicitly, trust. Empirical studies of friendship [Hays 1985, 1989; Oswald et al. 2004] point to the advantage of investing in fewer, more intimate relationships. When collaborations persist in groups, mutual trust will develop between individuals, leading to reliance on each other for help, emotional support and companionship [Baumeister and Leary 1995]. Several studies on friendship have established the association between interaction frequency, intimacy or trust and the strength of social relationships [Larzelere and Huston 1980; Rotenberg and Morgan, 1995]. Hays [1989] compared close and casual friendships and found that close friendships showed more interactions during the week, across a wider range of days, times and locations, than casual friendships. In terms of benefits received, close friends offered more emotional and informational support than casual ones. Indeed, the most commonly used classification in these studies is best, close and casual friend [Rose and Serafica, 1986; Oswald et al. 2004]. Oswald et al. [2004] assessed friendship in terms of different dimensions of trust-oriented behaviours (positivity, supportiveness, openness and interaction), showing a clear distinction between higher-scoring best friends on the one hand and lower-scoring close and casual friends on the other. Rose and Serafica [1986] found that casual friends required less emotional involvement than best or close friends, but

were more dependent on opportunities for contact. In contrast, best friendships were seen as being self-maintained.

Social preference based on reciprocity is the principal assumption underlying psychological theories of friendship such as exchange theory [Baumeister and Leary 1995; Brown and Brown 2006]. Empirical studies on the psychology of friendship suggest that layers of friends with increasing intimacy form the structure of personal ego networks with typically 10-12 close friends and a larger number (120-150) of casual friends or acquaintances [Hays 1989; Wellman et al. 2006]. Alternatively, social structures might arise from the dynamics of interaction as proposed by Heider's balance theory [Heider 1946], in which social networks are determined by the balance between positive relationships between friends and antagonistic relationships between enemies. Whitmeyer [2002] adapted Heider's balance theory's constraint that too many connected friends increases the probability of conflict, to develop a social network model that predicted that fewer more intense relationships (circa 5) would emerge at the expense of more diffuse ties. While the empirical literature agrees that the complex structure observed in human groups is composed of layers of relationships with increasing levels of intimacy [Baumeister and Leary 1995; Hays 1985, 1989; Wellman et al. 2006], there is less consensus about the processes underlying the emergence of these layers.

2.2. Computational Modelling of Trust

Computer simulations of repeated prisoner's dilemma games have demonstrated that cooperative strategies will spread in populations when histories of interactions are accessible [Nowak and Sigmund 2005]. The antecedents for trusting relationships in cooperative behaviour and reciprocal exchanges were investigated by Nowak and Sigmund's [1998, 2005] computational modelling of reputation ('image scoring'), showing that cooperation and altruism are likely to be widely adopted in populations where reputations (i.e. the image score) are publicly visible. Hardy and Van Vugt [2006] also proposed that reputation systems are a necessary prerequisite of evolutionarily stable cooperation in large groups; furthermore, Roberts and Renwick [2003] demonstrated, in both experimental studies and computer simulations, that individual reputations based on histories of collaboration lead to formation of social relationships. Hruschka and Henrich's [2006] model produced emergent social structures and preferential social relationships when agents possessed strategies favouring known collaborators, and social preference strategies ('cliquers') spread in populations across a range of cost/benefit ratios and defect rates. However, their model depended on a memory constraint, so only a limited number (10) of stronger relationships could be developed. These simulations of cooperative behaviour do not account for the development of trust in social relationships that apply to observed human social structures.

Trust models are very diverse, ranging over the object of trust (person/product), concepts of trust and mistrust held by trustor and trustee roles [Castelfranchi and Falcone 2010] and the decision options, agents' reputation, visibility and context [Sutcliffe 2006]. Models may be based on ratings or transactional views of trust which depend on the ego's assessment of alters' actions or message content in e-mail [Golbeck and Hendler 2006]; however, the transaction view does not account for longer-term development of trust-based relationships. Formal socio-cognitive models of trust [Falcone and Castelfranchi 2001a, 2001b] evaluate the various factors that influence trusting relationships, such as the degree of delegation between the two parties, the motivations, risks and goals shared by the parties to establish the need for a relationship, and properties which can be evaluated to establish their reputations. The socio-cognitive theory of trust [Falcone and Castelfranchi 2001a, 2001b; Castelfranchi and Falcone 2010] represents a considerable synthesis of the

literature; however, the equations and premises contain many assumptions which are not based on empirical evidence. Furthermore, the predictions of the model have not been validated in case studies or experiments.

In a more organisationally oriented model, Riegelsberger, Sasse and McCarthy [2003] analysed trusting relationships and trust-enforcing institutions and proposed mechanisms for promoting trust in communities, with design guidelines to facilitate evaluation of intrinsic trust of the reputation of the trustee, as well as feedback from experience.

The concept of trust has been elaborated into credibility in studies of websites [Fogg et al. 2001], where trust is an attribute assigned to websites based on the user's assessment of several factors such as usability, accuracy of information, brand and persuasive features. However, Fogg proposed no specific model of trust *per se*. A more elaborate model based on qualitative and quantitative research is the two-process model of trust in which first impressions, based on the interface look and feel, are distinguished from more detailed evaluations, based on the analysis of source credibility, personalisation and predictability [Briggs et al. 2002].

3. MODELLING SOCIAL RELATIONSHIPS

This section summarises the behavioural predispositions that might lead to the emergence of complex social structures. A more extensive treatment is given in Sutcliffe et al. 2012]. The conceptual model presented in this section builds on the relational view of trust as found in the OED definition, "Firm belief in the reliability, truth, or ability of someone or something: relations have to be built on trust". The main tenets of the model are that trust is derived from mutual, positive, reciprocated interactions between individuals, and that the frequency of those interactions leads to the development of trust and hence the increasing strength of social relationships. The model also proposes mechanisms governing the emergence of social networks from behavioural predispositions to concentrate social interaction on a few partners rather than spreading interaction more evenly among many individuals. We propose that trust underpins the formation and maintenance of social relationships and that the degree of trust in social relationships is a function of the frequency of positive, cooperative interaction.

3.1 Role of Trust in Relationship Formation

The relationship between interaction frequency and relationship strength was established by Hays [1989], who compared close and casual friendships and found that close friendships showed more interactions than casual friendships. Furthermore, the frequency of contact, both face-to-face and by electronic communication, has been correlated with rating of emotional closeness among friends [Roberts and Dunbar 2011a; Rose and Serafica 1986; Wellman et al. 2006]; and when contact frequencies decreased due to separation, relationship closeness also declined [Roberts and Dunbar 2011b, 2012]. Therefore, there are reasonable grounds for proposing a model of relationship formation based on interaction frequency. However, relationship closeness is not a linear function of interaction frequency; instead, intimacy reaches a plateau where the relationship becomes affect-based, following the change from calculative to relational-mode trust in social relationships [Rempel et al. 1985; Rose and Serafica 1986; Sutcliffe and Wang 2012]. Affect-modulated social interaction accords with theories of trust in social relationships [Rempel et al. 1985; Ostrom 2002]; moreover, affect stabilises and intensifies relationships in pair bonds and mother-infant relationships [Bowlby 1990; Hinde and Stevenson-Hinde 1990].

If individuals interact socially and collaborate over an extended period of time, the level of mutual trust should increase to the stage where emotion (i.e. pleasure in the

alter's company) becomes more important than the rewards of collaboration. The interaction between trust and relationships is summarised in Figure 1, showing three hypothetical scenarios of trust formation for each SBH relationship type, assuming different distributions of social interaction. The upper line (strong ties) shows trust developing to a plateau as the law of diminishing returns applies. Crosses indicate non-cooperative behaviour by the alter agent. Strong ties are relatively immune to such events, modelling the intuition of forgiving the transgression of best friends.

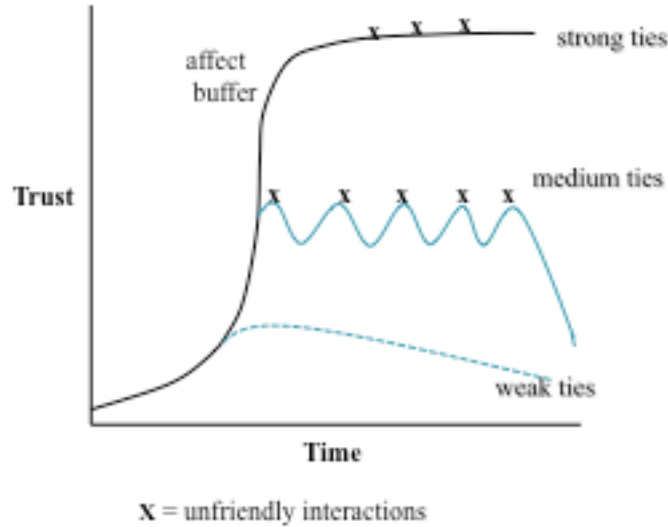


Fig. 1. Development and decay of trust over time, illustrating strong, medium and weak tie scenarios with different rates of interaction and decay from neglect and/or non-cooperative responses [see also Sutcliffe et al. 2012]

The strong ties trajectory assumes that many social interactions in a relatively short time period increases trust and intimacy. Intimate relationships pass through an affect threshold whereby relationship maintenance is governed by emotion/pleasure in the alter's company; hence, less cognitive resource needs to be devoted to relationship maintenance. These strong ties may be more resistant to non-cooperative responses, reflecting the intuition that best friends/partners are given the benefit of the doubt when interpreting non-cooperative responses. However, if non-cooperative events from the alter persist, or no cooperative interactions reinforce the relationship, even strong ties will decay.

In the medium ties scenario, there are fewer cooperative interactions, or interactions are spread over a longer time period; hence trust develops at a lower level and does not reach the affect threshold. Medium ties will therefore be more susceptible to defect responses, illustrated by the sine wave time-course showing conflicting influences of cooperative and non-cooperative behaviour by the alter. As with strong ties, if interaction ceases or non-cooperative responses continue, the relationship will wane. Finally, weak ties assume a low level of interaction, and similar susceptibility to negative interactions; however, given the low level of trust the effect of lack of reinforcement or non-cooperative response is termination of the relationship.

The number and intimacy level of relationships should be governed by the distribution of interaction among alters and constraints on an individual's social time budget. More time has to be devoted to maintaining strong-tie relationships to keep them committed, and this leaves a finite resource for other relationships. While medium- and weak-tie relationships can be maintained by lower interaction

frequencies, over-dilution of social attention among alters runs the risk of losing trust by favouring others and exacerbating the cognitive load of social management [Sutcliffe et al. 2012].

3.2 Rewards/Fitness Drivers in Social Relationships

Close relationships offer important benefits in terms of stress reduction and social support: proximity of others who can be trusted to come to one's aid should reduce stress by their reassuring presence and the promise of intervention in conflicts by active support, as has been documented in primates [Crockford et al. 2008; Dunbar 2010]. Among humans, there are striking effects of network membership on wellbeing. Fowler and Christakis [2008], for example, have shown how the state of happiness of one's friends and friends-of-friends can influence one's own happiness. Nonetheless, even though large support groups will always be more effective than smaller ones, there must inevitably be a trade-off between the benefits that accrue and the number of alters who provide that benefit. This is likely to be so for two reasons.

First, only a few individuals may be able to provide the benefit at any one time, since many individuals may interfere with each other (a too-many-cooks-spoil-the-broth effect). Second, close relationships are usually reciprocal [Brown and Brown 2009], so while each relationship accumulates potential benefit to the ego, it does so at the cost of exposing the ego to the risk of being called upon to reciprocate commitments to the alters. A cost-benefit trade-off may limit investment in a small subset of individuals (special friends) rather than spreading one's social time budget evenly among more individuals. We argue that costs of reciprocating support, i.e. risk, increase linearly with more very close relationships; however, benefits produce diminishing returns since special friends need to be in close proximity to help. These special friends are the most likely source of costly emotional and instrumental support, especially in times of emergency. Time investment costs of reciprocal commitment may also rapidly outweigh the benefits of emotional support, to limit the optimal size of the support clique to a small number.

Investing in more stronger relationships could improve the pool of individuals from which alliances with different functions might be formed [Dunbar 2010]. Alliances, we argue, reach a law-of-diminishing-returns plateau once a certain number of alliance partners has been reached, such that there are always sufficient partners available to choose from and any further increase results in more stress in managing relationships [Heider, 1946] and an increasing theory-of-mind [Mitchell 1997] memory burden in tracking more complex and dynamic relationships.

If these proposed trade-offs exist, then they should be manifest in social time and relationship-management strategies; for example, to favour the few at the expense of the many, and to divide the ego's social time accordingly. However, as each layer of relationship intimacy conveys a different advantage to the individual so that an individual's social time budget is spread over many individuals [Dunbar et al. 2009; Roberts and Dunbar 2011a], a skewed distribution is produced with a few intense relationships and many weaker ones. The model can be interpreted in two time perspectives: first within individuals' lifetimes when predispositions might be learned from experience, and secondly in evolutionary time, across generations where behavioural predispositions have a genetic basis and therefore become selected for over longer time periods. For the Internet, the behaviour time perspective is more pertinent, although the conclusions from the simulation apply to evolutionary time as well. The evolutionary simulation results may also be interpreted in memetic terms [Dawkins 1976] as competitions between behavioural norms for social preferences, work/social time balance and cooperative/non-cooperative behaviour, given perceived rewards for wellbeing and costs.

This leads to the following hypotheses concerning the predispositions that might arise to create a social structure with layers of relationships at different levels of intensity:

- 1) Social relationships at different levels of intensity will emerge from the predisposition to 'favour-the-few' in social interaction.
- 2) A layered social structure will only emerge if social interaction is frequent, i.e. a tendency to spend more time in social interaction at the expense of other activities, e.g. work and other solo non-interactive behaviour.
- 3) A few intimate social relationships will improve wellbeing more than many less intimate relationships.
- 4) A small number of close social relationships will provide sufficient alliances to improve social wellbeing and reduce stress.
- 5) Many weaker relationships will improve networking and information access which could produce both social and material rewards. This conjecture is based on Granovetter's [1973] view of the value of weak ties.

These hypotheses associate fitness criteria (in the evolutionary perspective) or rewards (in social learning, wellbeing, alliances) related to the costs (risk, stress) which accrue at each relationship level, with a further benefit of resource to reward time spent in work (non-social activity). The hypotheses will be tested by systematically varying these fitness criteria in a computer simulation of social interaction. Other analyses relate to cost-benefit trade-off between the time spent socialising or working at each level of relationship intensity.

The behavioural model of social interaction is based on the following assumptions:

- (i) Increase in relationship strength is related to the frequency of social interactions, although the incremental increase is subject to a law of diminishing returns. Stronger relationships increase at a slower rate, hence closer relationships are less susceptible to change and cost less to maintain.
- (ii) Relationship strength wanes over time at a slow rate. This models the empirical observation that emotional closeness in relationships declines unless it is maintained by frequent social interaction [Roberts and Dunbar 2012; Wellman et al. 2006].
- (iii) People interact with others following various strategies, e.g. either favouring individuals with whom they have already frequently interacted with (favour-the-few) or spreading their interaction among more individuals (favour-the-many).

The tendency to favour the alters you have interacted with most frequently is coupled with the affect-based modulation that amplifies this tendency as the relationship becomes more intense. This is a plausible behavioural mechanism for implementing the cost-benefit trade-off hypotheses we proposed in Figure 1, without explicitly having to count or estimate the costs and benefits of each social interaction, which would be unnecessarily cognitively demanding.

The computer simulation, therefore, is intended to be a unified account of social behaviour which may arise from learning within an individual's lifetime and the evolution of social predispositions across generations.

4. MODEL DESIGN: SIMULATING SOCIAL RELATIONSHIPS

We modelled the strength of relationships as a function of interaction frequency. Frequent, positive and reciprocated interaction engender trust between individuals, and we argue that strong social relationships are associated with a high degree of trust [Brown and Brown 2004]. The frequency of social interaction is constrained by the time individuals devote to the activity, hence we modelled the competing demands between time spent in resource acquisition (i.e. work) and time spent socialising, which yields increased wellbeing, lower stress and other benefits of social

cooperation, including alliances and inter-group competition. For simplicity, we restricted modelling to a simple trade-off between spending more or less time working or socialising, reflected in the ratio of ‘work’ vs. ‘socialise’ (WS).

We used an agent-based model in which agents had choices: first whether to socialise or work, and then, when socialising, which agent to interact with. Agents initially meet randomly, but maintain a memory of alters they have met, so in subsequent turns they apply preferential strategies to constrain which alters they socialise with. The agent interaction model uses a smoothing algorithm to modulate both the effect of interaction frequency on relationship strength increase and decrease via waning, so stronger, relationships are more resilient than weaker relationships. Details of the algorithms are given in Sutcliffe and Wang [2012].

The main components of the model are:

- Agents interact and form trusting relationships. Frequency of interaction is equated with strength of relationships, following empirical observations [Roberts and Dunbar 2012; Rose and Serafica 1986].
- Responding agents (alters) may choose to accept or reject an invitation to interact, hence relationship strength accumulates for positive interactions, and is decreased by negative interactions.
- Agents have social preference strategies for initiating interaction according to the history of previously successful interactions. Four agent strategies are tested ranging from favour-the-few to favour-the-many.

Each focal agent (the ego) within the population is assigned a turn, so there is an equal opportunity to interact. Initially, the responding alter agent is selected at random. However, once a history of interaction develops, the ego’s choice of alter is governed by its strategy as follows:

Favour-the-few preference: the history-dependent search function favours cooperating with alters with whom the agent has had more previous successful interactions. This encourages development of strong ties; however, since the function is stochastic, there is a low probability of initiating interaction with low-trust alters and strangers without a previous interaction history. As relationships develop, the search function is progressively biased towards stronger ties.

Midway preference: the search function is biased towards the mid-point in the trust relationship distribution.

Favour-the-many: the search is biased towards the low trust/interaction frequency part of the distribution.

Staged: in this dynamic strategy, the search bias is changed over time (model cycles) so the ego agent initially favours the few but progressively favours initiating proportionately more interactions with strangers and low-trust (favour many) partners. This implements the behavioural predisposition manifest over an individual’s lifetime, that strong ties are formed earlier in life, while weaker ties accumulate in later life [Hays 1989].

Social mediating technology, such as Facebook, might influence (a) the number of accessible alters for social interaction, (b) the frequency of interaction, and (c) the quality increment in trust for each exchange [Boyd and Ellison 2007; Ellison and Boyd 2013; Sutcliffe et al. 2011]. The last influence might arise because face-to-face contact facilitates stronger emotional interaction through social presence and physical interaction, which are probably not as strong in the cyberworld even with virtual characters (e.g. SecondLife). Another possibility is that computer-mediated communication reduces social inhibition [Postmes et al. 2000] so people might be less cooperative especially in weaker social relationships.

The experimental parameters (independent variables) in the computational model are:

Defect rate: percentage of turns in which the respondent will not cooperate. This models a possible social disinhibition effect.

Waning rate: a parameter that decreases relationship strength each cycle. This parameter models the decline in trust. In the absence of social interaction, trust values will decline and relationship strength will decay from strong to weak to zero.

Number of cycles: cycles in each run. This models the effect of opportunity for frequent interaction afforded by social media. Since social media facilitate a potentially larger population to interact with, this parameter evaluates the interaction frequency.

Proportion of agents by strategy in the population: this reflects our assumptions about the initial population, starting with an even proportion of all strategies.

The model outputs (dependent variables) are:

Relationships categorised by trust strength between agents for each strategy.

Frequency of relationships for each agent ranked by relationship strength.

Average number of relationships per agent, divided into strong, medium and weak ties.

Since the frequency distributions of relationships by trust strength for each agent showed a power law pattern, we used a tercile (33-33-34%) split on the range of trust that divided the population into a few agents in the upper third, a few more in the mid-range, while most agents fell within the tail of a power law distribution. This rationale divided the continuous distribution into strong, medium and weak ties.

Details of the model and source code can be inspected in the ABM repository under the title Social Trust Model at <https://www.openabm.org/model/4458/version/1/view>

5. SIMULATION RESULTS

5.1 Validating the Trust Model

In previous work [Sutcliffe and Wang 2012] we investigated several interaction frequency-relationship intensity algorithms (+/- smoothing), with a range of waning and defection parameters, and social strategies to observe their effect on social structure. The results pertinent to the technology hypotheses are summarised in this section.

A key assumption in the trust model is that the relationship between cooperative interaction frequencies and trust is logarithmic rather than linear, reflecting the emotional damping effect illustrated in Figure 1. We tested this assumption in a 2 x 2 experiment varying log/linear trust increase and waning, to assess the effect on average trust levels for strong, medium and weak relationships. Simulations were run for 1000 to 5000 cycles, populations 200-500 with parameter ranges for linear increase and waning from 0.01 to 0.1; and log compression ratios from 1 to 10. A sample of the results for each simulation version is illustrated in Figure 2, which shows the average ties per agent divided into the tercile categories (strong, medium, weak and total) for runs with 200 agents. The pattern was similar for 300, 400 and 500 agents.

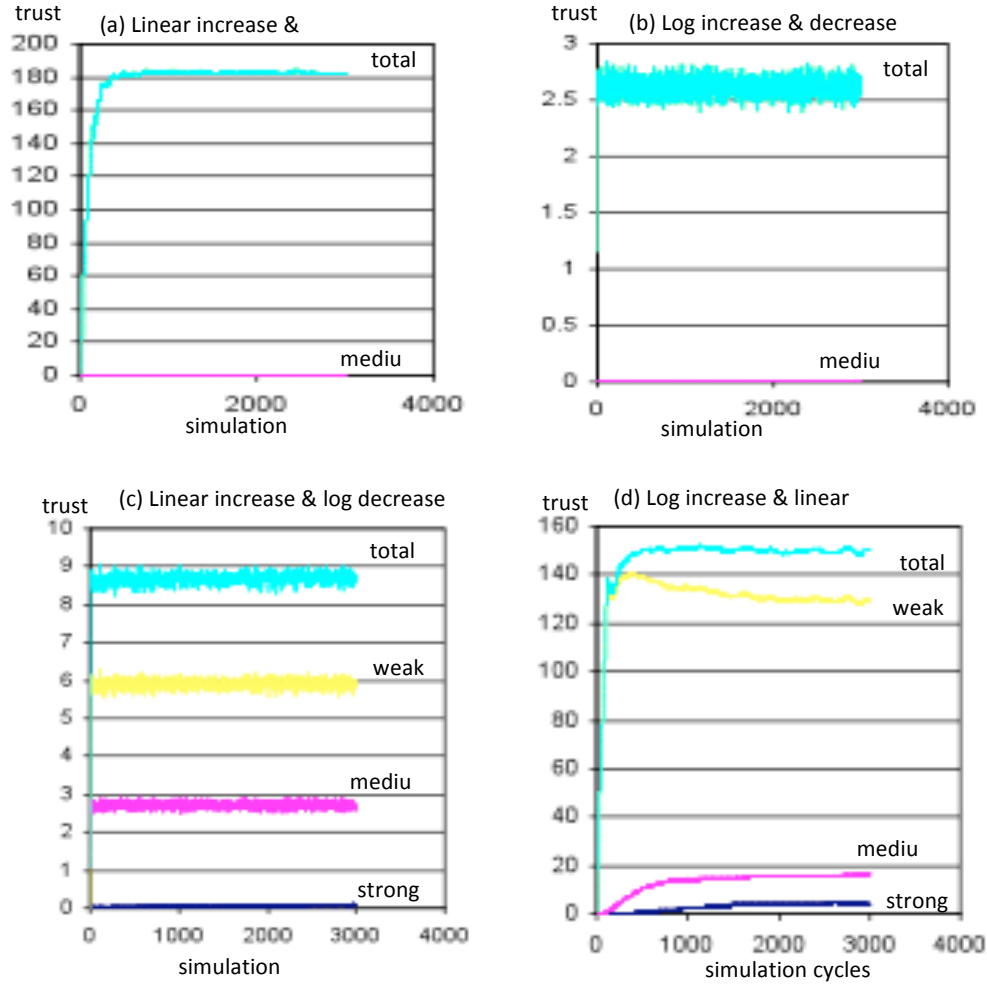


Fig. 2. Simulation results showing average (strong/medium/weak/total) ties per agent for each trust model version. No strong ties were present for (a) and (b) conditions so the total equates to weak ties.

Linear increase and decrease models produced many weak ties but no strong or medium ties when the increase rate was just high enough to overcome the waning rate, since many ties were initiated but waned before they could become strong. Increasing the difference between the trust gain and waning rate biased the result towards more stronger trust ties.

Log increase and decrease simulations produced a low number of overall ties (2-3) with no strong or medium ties. Log increase favoured the gain in trust; however, log waning rates were proportionately higher for weak ties, hence these tend to be extinguished and the simulation shows instability with oscillating numbers of weak ties.

Linear increase and log decrease produced about 8-9 ties overall with some differentiation of about 3 medium ties to 6 weak ties. Linear trust increase did not develop ties as quickly as the log version, while the log decrease effectively reduced all ties; hence the number of relationships remained too low.

Log increase and linear decrease. This version produced the best results for SBH with a balanced distribution of 130 weak ties, 14-16 medium ties and 5-6 strong ties. Therefore this model was adopted for further testing with the following parameters:

- Frequency-intensity algorithm: log increase/linear decrease.
- Smoothing interval on increase = 10%: the log algorithm is applied progressively on each 10% of the range of trust values (0-100).

- Waning rate = 0.08: total trust in each relationship is decreased by 8% per cycle, varied between 0.01 to 0.10.
- Population size: 200, 300, 400, 500.

The trust model was run with the above range of parameters to evaluate the social media assumptions concerning interaction quality and frequency. As the number of cycles is increased trust values for each alter relationship approach an asymptote, with the strength level being a function of the ego agent's strategy (i.e. favour-the-few relationships were stronger). This indicates that increasing the frequency of interaction beyond a certain threshold does not affect the distribution of social relationships in our model. Simulations with high trust gain and low waning decrement settings all produced high-trust relationships approaching 300 per agent with $N = 300$ agent populations after 3000 cycles. Higher waning rates and lower trust-gain rates slowed the rate of increase but the trend towards maximal strong-tie relationships continued. Conversely, when the decrement rate was lower than the gain, not surprisingly many relationships were formed and then extinguished; consequently the number of ties per ego and their strength decrease with high waning rates. Higher waning rates reduced the longevity and overall number of relationships. This suggests that computer-mediated-communication interactions may not be as effective as face-to-face in maintaining trust. The log/linear model produced distributions with an upper limit of circa 150 overall ties per ego within a range of population sizes from 200 to 500. In populations with different agent strategies, strong and staged produced levelled social structures; medium and weak did not. The simulation shows that the number of overall relationships appears to be a function of agents' partner-preference strategy, and the trust increase/waning model.

Two impacts of defection were tested to investigate the effect on the trustor (initiating agent) and 'mutual suspicion' when both parties become less trusting of each other:

- Asymmetric: (trustor only) if B defects then only A's trust in B is reduced
- Symmetric: if B defects then the relationship in both directions (A-B, B-A) is reduced, if a positive reciprocal value exists.

With the asymmetric model, defect rates of up to 10% had no effect on the overall average number of ties per ego for all strategies; rates above 10% progressively reduced the number of strong and medium ties for the strong and staged strategies, with a small decrease in overall (weak ties) for weak and medium strategies. The symmetric model produced a stronger but similar effect although at a lower rate of defects, 5-10%, as may be expected since both relationships are decremented.

Higher defection rates progressively reduced the number of strong and medium ties which decline to zero with a rate higher than 10%, so strong ties can not develop since high frequencies of defection prevent the development of trust. The trust model appears to be robust in the presence of modest defection rates up to 5%, although higher rates prevent strong and medium ties developing.

If interaction quality in social media is lower than face-to-face then interaction frequency will have to be increased to compensate. However, the plateau effect on trust in all stronger relationships suggests that increasing interaction will not compensate for maintaining trust, although it might help form more weaker relationships. Our model is sensitive to waning rate, which indicates that if trust in social relationships decays rapidly, then such relationships will be extinguished; however, most social relationships are based on strong trust which does not decay rapidly [Brown and Brown 2006]. The sensitivity analysis described above demonstrated the model was robust using a wide range of parameters; but, as expected, higher waning reduced the number of ties for all agent strategies.

5.2 Investigating the Emergence of Social Structure

5.2.1 Social Structure Simulation

We investigated how well each social preference strategy would survive, given fitness criteria in terms of wellbeing as the benefit of social interaction and close relationships, and resource acquisition as the benefit accruing from foraging and finding food. Social rewards were set against costs by defining five fitness criteria to model the support and sympathy group trade-offs. Three criteria made positive contributions to fitness (wellbeing, alliances and resource), while two (risk and stress) had adverse effects. Wellbeing was a function of relationship trust and interaction frequency, alliance was calculated from relationship trust and total relationships, while risk and stress were governed by the number of interactions and total relationships.

Wellbeing(WB) for the each individual agent was calculated as:

$$WB = MT * (I_{in}/I_{out})$$

where

MT is the mean trust for the individual agent (ego), i.e. total trust in the ego's relationships/N relationships. Trust is a function of interaction frequency.

I_{in} is the total frequency of interactions initiated by other (alter) agents with the ego during a lifetime

I_{out} is the total frequency of interactions initiated by the ego with other (alter) agents during a lifetime.

Hence wellbeing is modulated by the balance of interactions, with individuals receiving more attention from others having better wellbeing than net donors of interaction

Alliance (AL) was defined as

$$AL = MT * (a * N)$$

where

MT is the mean trust defined for WB, N is the total number of the ego's relationships at the end of each lifetime cycle and (a) is a coefficient that moderates the influence of total ties, set to $a = 0.05$. Agents with more relationships and higher trust are favoured, although the tie volume effect is reduced by the coefficient to reflect the importance of stronger relationships.

Resource (R) was a function of work turns:

$$R = F_r * N$$

where

F_r is the total frequency of work turns by the ego agent in a lifetime cycle

N is the total number of the ego's relationships at the end of each lifetime cycle

Resource benefits accumulate from work turns augmented by the number of ties per ego to model the benefit of relationships with others providing information which could subsequently improve the individual's performance. This function favoured agents with more, weaker ties. Fitness is driven by the basic need to feed and find food, so there will always be an upper bound on the amount of time that any species can afford to devote to social interaction [Dunbar 1992; Dunbar et al. 2009]. Hence we simulated partitioning of social/work time in the range of 1W: 1S, to 8W:1S which reflects the time distribution in observed in humans and other highly social species.

Risk (RS):

$$RS = MT / (b * I_{tot})$$

where

MT is the mean trust as defined before, I_{tot} is the total frequency of interactions (inbound and outbound) with the ego during the lifetime and (b) is a coefficient

that moderates the influence of total interactions, set to 0.005. More interactions increase the individual's risk exposure from injury incurred is aggressive interactions.

Stress (ST):

$$ST = N$$

where

N is the number of total ties for ego *i*. Stress is a simple function of the number of relationships that an individual has. Managing relationships and social living induces stress as well as having benefits.

Fitness criteria for all agents were evaluated and a net fitness determined after a lifetime of 3000 interaction cycles. Agents' scores on each criterion were ranked; the ranks were modulated by the weightings selected for each simulation run, then summed to create an aggregate fitness, as follows,

where, R = resources, WB = wellbeing, AL = alliances, RS = risk and ST = stress:

$$\text{Fitness} = (R + WB + AL) - (RS + ST).$$

Selection eliminated the weaker 20% of the population, ranked by fitness values. Breeding replaced the agents by replicating individuals drawn at random from the top 20%. The initial population was set at N = 300 with equal numbers of the four strategies: strong, medium, weak ties and staged.

All five fitness criteria were tested (WB, R, AL, RS, ST) with runs covering all combinations of weights from 1 to 5 and then with an extended range of weights up to 10 to control for population scaling effects. The effect of work/social interaction ratio of turns and the rate of defections on selected results from the fitness and agent strategy selection analysis were also explored. W/S ratios were varied between 8:1 and 1:1, covering reasonable assumptions about the amount of time people can devote to socialising. Given an average time awake of 16 hours/day and assuming an 8-hour working day, since humans are a highly social species a split of 50-50 social/work time is reasonable. However, some people engage in longer work hours and spend little time socialising during their leisure, so the percentage of social time can be low; hence we modelled a range from high to modest social interaction. We were interested in reflecting a range behaviour, therefore lonely individuals who spend little time socialising were represented in ratios favouring work. Defect rates from 1% to 0.25% of turns were selected based on preliminary testing of the trust model [Sutcliffe & Wang 2012] which showed that higher defect rates prevented the development of trust for all stronger relationships while lower defect rates had little effect on simulation results.

Simulations were run for 50 generations with 3000 cycles/generation, to produce outputs showing the populations of agent strategies across generations and average relationship strength divided into strong, medium and weak tie ranges. The parameter settings are summarised in Table 1.

Table 1. Summary of experimental parameters

Parameter	Setting/range
Fitness criteria	1..5
W/S turns	8:1, 5:1, 4:1, 1:1
Defect rates	1% to 0.25%
Agent population	300
Seed population	25% agent /strategy, 4 strategies
Cycles/generation	3000
Generations/simulation	50
Selection rate	20% elite

5.2.2 Social Structure Results

This section reports, first, a sensitivity analysis of different reward conditions to discover which social structures emerge under a wide range of fitness criteria weightings. The relationship patterns of social structures are reported and related to the fitness parameter space. The average frequency of relationships was output by trust strength for all agents produced from each simulation run after 50 generations. Given the large number of simulations ($5^5 \cdot 25 = 3125$), cluster analysis was carried out on the results to aggregate runs by similarity in the overall number of ties, and the tie distribution. Two-level K-means clustering algorithms were used to find the optimal number of clusters with the smallest standard deviation for each cluster. The fitness criteria weightings for all the runs in a pattern were averaged to create a profile of where each pattern lay within the range of all combinations of the fitness criteria.

The second set of results reports a further simulation into the competitiveness of the two key agent strategies (staged and strong) which delivered levelled social structures. These simulations tested whether these strategies would spread from a very small baseline in a population under a range of fitness criteria.

Relationship Patterns and Fitness Criteria

The five clusters are defined by the similarity of their output relationships (strong/medium/weak) patterns as shown in Table 2. Inspection of a sample of the results indicated that the dominant pattern (see Figure 3) produced many weak ties/ego, while a minority of runs produced more strong and medium ties. Social networks with an average size of 120 alters, which were nearly all weak ties, was the most frequent pattern, accounting for 51.9% of the runs.

Table 2. Cluster analysis summary showing total number of runs for each pattern and average and standard deviation () of ties for each pattern

Pattern	Total	Average ties (Standard deviation)			
		Strong	Medium	Weak	Total
Many weak ties	436	0.02(0.23)	0.10(0.91)	120.61(1.03)	120.73(1.00)
Medium+ weak ties	2504	1.02(0.14)	6.91(0.40)	104.66(0.87)	112.59(0.99)
Many strong ties	155	7.90(0.59)	20.83(1.12)	134.29(1.80)	163.03(3.36)
SBH	29	5.42(0.96)	15.98(1.84)	125.88(3.24)	147.28(5.94)
Only weak ties	1	0	0.00	183	183

The second most frequent pattern (44.5%) was a ‘small core’ network with some 1 strong and 7 medium ties, and an overall network of around 112 alters. The third most frequent pattern (2.6%) had about 8 strong ties, 21 medium ties and an overall network around 160 individuals. This was followed by SBH¹-compliant networks (0.6%) with an average of 5 strong ties, 9 medium ties and a network size of 147. The final pattern had no strong ties, very few medium ties and an overall group size around 196.

¹ SBH =strong ties in the range 4-7, medium ties 8-12, weak ties 130-160.

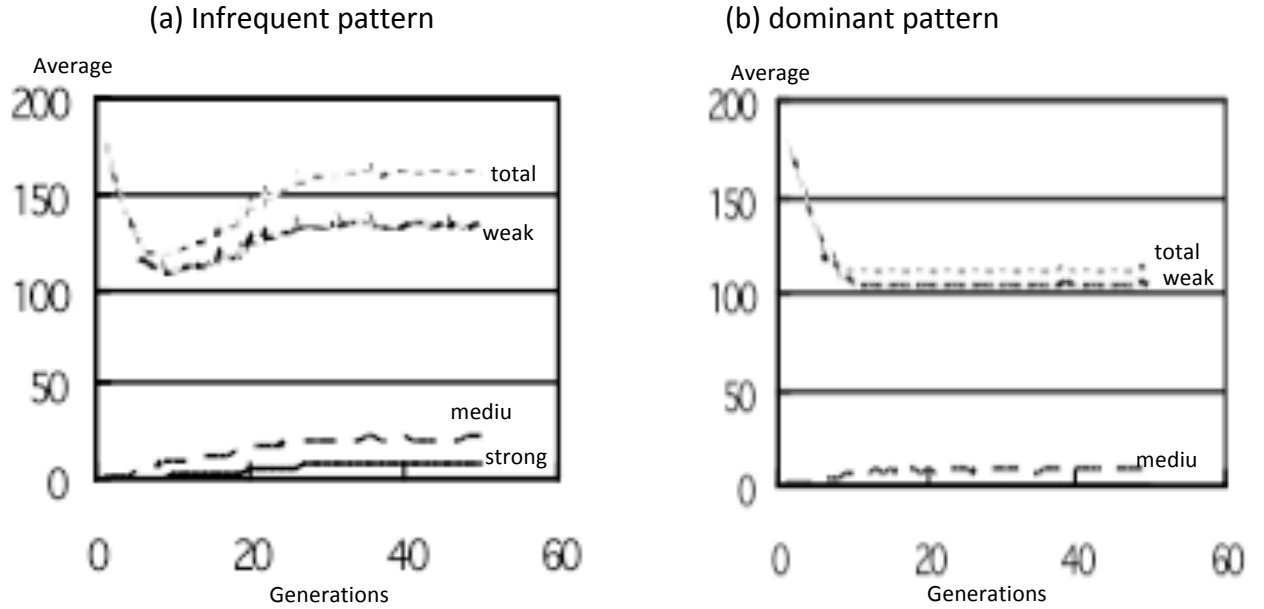


Fig. 3. Sample results showing the mean frequency of ties/agent over 50 generations for (a) the less frequent and (b) the dominant pattern

Table 3 shows the weightings for each strategy cluster on the five fitness criteria (resources, wellbeing, alliances, risk and stress) for each of the five patterns which emerged from the simulations. The percentage in column 1 is taken from the total runs/pattern in Table 2.

Table 3. Mean fitness weightings for each social structure pattern

Pattern	% total runs	Mean weighting (Standard deviation)				
		Resources	Wellbeing	Alliance	Risk	Stress
Many weak ties	13.95	3.22(1.40)	1.79(0.93)	1.69(0.86)	4.31(0.85)	2.77(1.40)
Medium+weak ties	80.128	3.07(1.40)	3.16(1.38)	3.13(1.36)	2.87(1.37)	3.13(1.40)
Many strong ties	4.96	1.52(0.72)	3.68(1.32)	4.41(0.80)	1.68(0.92)	1.68(0.88)
SBH	0.0928	1.90(1.05)	3.72(1.25)	4.24(1.09)	1.83(0.97)	2.14(1.22)
Only weak ties	0.0000	1.12(0)	1.62(0)	4.00(0)	4.75(0)	1.00(0)

The weak-ties and medium-/weak-ties patterns, which accounted for 96.5 % of all the results, had an even distribution of weightings for all fitness criteria in the range 2-3.5, apart from risk, with the difference between the two patterns appearing in weightings for risk and stress, with higher resource and lower alliance weights for weak ties and the converse for the medium/weak ties. The only weak-ties pattern was so infrequent that it can be discounted. There were only minor differences between the SBH and strong-ties patterns in reward criteria weightings: slightly higher wellbeing, resource and stress produced the SBH pattern, whereas a higher alliance weighting produced a stronger-ties pattern. Risk and stress were low for both of these patterns.

The fitness criteria weightings for each pattern were clustered to map the distribution of the patterns within the overall space of fitness weightings. The best resolution was produced by 5 clusters for all the patterns and 4 for SBH (for which a 5-cluster solution produced one group with just one member). For the two most frequent patterns (weak ties and medium + weak ties), the five clusters were:

- High stress, risk and resources
- High risk, moderate stress and resources
- High risk

- (iv) High resources and risk
- (v) High risk and stress

where high values are weightings 4 and 5 and moderate = 3, with all other values (1 and 2) classed as low.

Exposure to high levels of risk and/or stress appears to prevent stronger ties from forming. The weak-ties pattern dominated for most fitness criteria combinations, i.e. for all runs except where wellbeing and alliance were high (4-5) and other fitness criteria were low (1-2), suggesting that social structures with many weak ties will emerge under a wide range of fitness conditions. However, the SBH-compliant and the stronger-ties patterns were both rare, and only emerged when social rewards were strong. The locus of the different patterns is summarised in Figure 4. SBH and the strong-ties pattern occupied a relatively small part of the overall fitness environment where wellbeing and alliance criteria were favoured and other criteria had low weights. The rest of the space is occupied by the weak or medium-/weak-ties patterns.

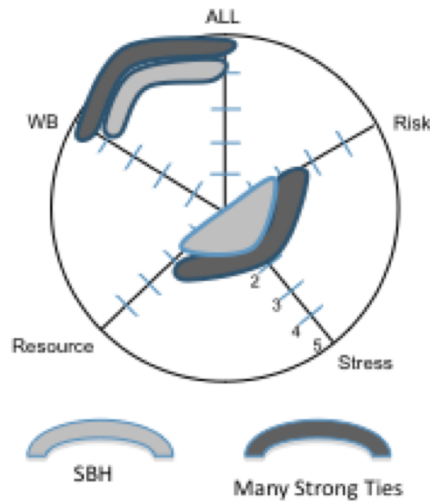


Fig. 4. Locus of the SBH and many strong-ties patterns in the space of fitness criteria weights. Dark shading indicates the SBH pattern, lighter shading many strong ties. All = Alliance, WB = Wellbeing fitness criteria. Weights increase towards the circumference of the circle.

Cluster analysis was also applied to runs which produced the SBH patterns to investigate possible groupings of fitness criteria. The average weights for the criteria in each cluster of SBH runs are illustrated in Table 4. High alliance weights appeared in three clusters. In cluster 1, alliance and wellbeing as well as stress were high; cluster 2 showed high alliance with modest resources but low wellbeing; while cluster 3 had both high wellbeing and alliance weights. In a minority of simulations (15%) high stress may occur when it is counterbalanced by high wellbeing and alliance rewards. Overall, an SBH social structure emerges when either alliances or wellbeing contribute strongly to overall fitness, and the disadvantage of risk and resources rewards are moderate. The alliance fitness function appears to be slightly more influential than wellbeing in determining SBH patterns, showing a high average weighting in three out of the four SBH clusters.

Table 4. Average weights of fitness criteria for clusters producing the SBH pattern

Cluster	Resource	Wellbeing	Alliance	Risk	Stress	Percentage
1	1.0	3.0	4.6	1.0	3.9	15
2	2.4	1.8	4.3	1.4	1.0	25
3	1.6	4.3	4.8	2.5	2.0	30
4	1.6	4.8	3.1	1.1	1.3	30

The strong-ties pattern occupied a similar distribution to SBH, with high wellbeing and alliance in four clusters, two of which had low weightings for risk or stress and one a moderate resources contribution. One cluster showed high alliance weighting with moderate wellbeing. Generally, the weightings of stress and risk in the strong ties pattern were lower than in SBH patterns.

WS Ratio and Agent Strategies

Inspection of the agent populations from the different patterns showed that the strong and staged strategies appeared in SBH, many strong and medium-weak patterns, while the weak ties results were dominated by the weak-tie pattern. To investigate the competition between social strategies, we simulated a population of 300 agents with four strategies, four preferences between social/work time; and four levels of social cooperation:

- Social preference strategy: strong, medium, weak tie preferences and staged as defined previously
- Social time predisposition, indexed as work:social interaction ratio with four settings <8:1, 5:1, 4:1, 1:1>
- Social cooperativeness:defect ratio for the alter in social interactions, four settings <0.01, 0.025, 0.05, 0.075>

50 simulations were run with a range of fitness criteria randomly selected from previous results that had produced the weak, medium-weak, SBH and many strong patterns in the previous results, e.g. SBH (1:4:5:2:1), weak ties (4:1:2:2:3), etc. The seed population (N = 300 agents) had a hierarchical distribution, containing an equal proportion of agents by strategy and then within each strategy an equal number of individuals for each WS genotype, and finally within each strategy/WS variant an equal number of individuals for each social cooperation level.

Fitness was calculated as before with elite 20% selection and breeding with both crossover and mutation. Crossover swapped strategies between agents (e.g. before, crossover agent A = strong and B = medium; after, crossover agent A = medium and B = strong). Mutation changed parameter settings in the defection rate and the WS ratio, which modelled the agent choice in the social/work time budget. The crossover rate was 0.9, i.e. 90% of the agents were selected per breeding cycle, and pairing was random. Mutation was random on pseudo-continuous distributions for:

WS (work/social turns) range <0.125..0.5>, (i.e. 8:1..1:1 ratios) divided into 1000 intervals

Defect range <0.01..0.075>, divided into 1000 intervals
with a rate of 0.01 applied to all agents, so 1% of the agents had a mutation of +/- one interval on both WS and defect parameters.

50 simulations were run, and the proportions of surviving agents after 50 generations are shown in Table 5. In the SBH-compliant, strong tie and weak + medium tie patterns, the staged and strong tie strategies dominated populations, while medium- and weak-tie strategies dominated only in the weak-ties pattern.

Table 5. Average number of agents by strategy for each social pattern emerging within the range of fitness criteria

	Staged	Strong	Medium	Weak	WS ratio
Weak ties only	2.18	1.62	273.62	22.58	0.499

Medium + weak ties	149.39	149.58	0.63	0.40	0.498
Strong ties core	138.38	161.55	0.03	0.05	0.152
SBH	134.34	165.66	0.00	0.00	0.266

Dominance of strong and staged strategies in simulations which produced SBH patterns with few strong and medium ties suggests that the favour-the-few strategy competes and overcomes defection, which decreased to nearly zero in spite of defecting agents being rewarded with extra work turns and hence resources. The WS (social time budget) ratio also changed to favour more work turns.

WS ratios stabilised at 0.26 and 0.15 in the SBH and strong-tie patterns respectively (i.e. more social interaction was favoured with ratios of 4-5:1), whereas in the two dominant weak, medium-/weak-tie patterns the ratio stabilised around 0.49, or an even ratio of turns. This suggests that medium and weak agents need more social turns to maintain even weak-tie relationships, whereas strong and staged strategies can maintain strong ties even with fewer social interactions. The strong and staged strategies dominated the populations for all patterns apart from weak ties, where medium and weak were more prevalent. The staged strategy had a slight advantage over strong for the weak only and weak+medium patterns, which might be expected from this hybrid strategy. Strategies which formed stronger ties were favoured even when criteria did not reward them in the medium-weak ties pattern; however, selection on the WS ratio toward 1:1 in this pattern (see Table 5) may have favoured survival of the strong ties strategies.

Since the strong and staged strategies were competitive over a wide range of fitness criteria, we tested their ability to spread in a population of weak-tie agents. When populations were seeded with 1% strong or 1% staged agents in a population of weak-tie agents with WS and defect gene settings assigned at random using a roulette algorithm, both strong and staged strategies spread rapidly to dominate the population within 10-20 generations. The model outputs (see Figure 5) were the same as those found in equal seed populations, with SBH-conforming patterns being produced under a range of previously observed fitness criteria weightings. The WS ratio stabilised at 0.15 while the defect rate was driven down towards the minimal setting with an average 0.05.

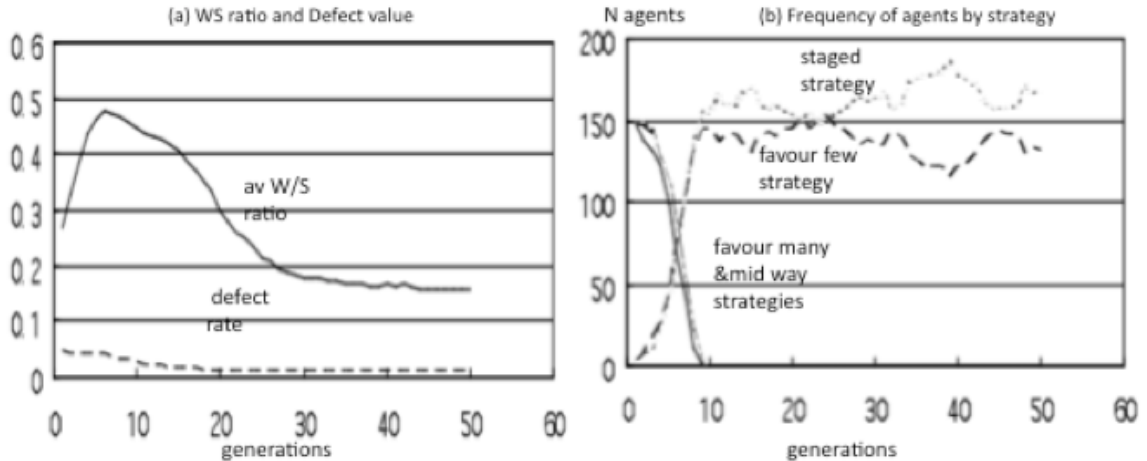


Fig. 5. Model outputs with 1% seed populations of strong and staged agents, (a) WS and defect rates, (b) average frequency of agents/ strategy.

Effect of Population Size and Selection Rates

To test the consistency of the results, further simulation runs were conducted with populations of $N = 500$. The trust increase smoothing and waning rate parameters were adjusted to reflect the larger population and hence decrease the frequency of

inter-agent interaction, so the strength of the trust-formation smoothing ratio (CR) was increased to 20, with a proportional increase in the waning rate to 0.18. The same result was produced with 0.5% seed populations of staged and strong agents, so it appears these strategies have a strong competitive advantage over a wide range of fitness criteria.

However, the larger population simulations did not produce either an SBH or a strong-tie pattern. Instead, the results were dominated by the weak-ties and weak + medium-ties patterns as found in the smaller ($N = 300$) population. Since the decreased probability of socialising may have reduced the persistence of strong relationships, further simulations were run extending the parameter range for the fitness criteria from $\langle 1..5 \rangle$ to $\langle 1..10 \rangle$. Increasing the fitness criteria range produced SBH-compliant and strong-ties patterns with similar clusters to those observed in the $N = 300$ population, but with increased wellbeing and alliance weights at 7-8, and low resources, risk and stress weights. The simulation model therefore scales to large populations, given appropriate adjustment of parameters to reflect the fact that less frequent interactions are inevitable within larger populations.

Since selection at 20% represents a severe mortality rate, simulations were also run with lower selection rates of 5% and 10%. These reduced the number of SBH patterns in the same fitness space, with only one SBH result being produced at 10% mortality (weighting 2-3-5-1-3 R/WB/AL/RS/ST) and none at all at 5%. Populations were dominated by weak-ties and medium + weak-ties patterns with a WS ratio close to 1:1. This suggests that emergence of SBH is influenced by both the selection rate and population size.

6 DISCUSSION

The trust model we propose simulated the emergence of social relationships which show a close fit with the empirical data on friendship and the role of trust in mediating relationship strength [Binder et al. 2012; Roberts and Dunbar 2011b, 2012]. The log/linear model simulated our proposed affect-buffered pattern for trust development. Since this model shows the emergence of social relationships of varying strengths under a wide range of reward conditions, it appears to be a suitable account of the cognitive mechanisms underpinning trust in human social relationships.

The trust model simulation generated social structures aligned with SBH's predictions for the numbers of relationships in each intimacy layer (5 support, 15 sympathy group, etc.). However, as the boundaries between layers of closeness in friendship are open to slightly different interpretations, e.g. 10 close friends v. 20-30 casual friends [Brown and Brown 2009; Hays 1989; Wellman et al. 2006] in contrast to the 5-15 support/sympathy groups predicted by SBH [Dunbar 2010], it is notable that the second strong-ties relationship pattern generated by the trust model fits with the former predictions. Hence the affect-modulated trust model fits with nearly all the empirical evidence and theoretical perspectives on human social relationships. Hypothesis 1 is supported from the observation that the favour-the-few strategy dominated all the simulations leading to the SBH pattern. However, hypothesis 2 (frequent social interaction) needs to be qualified since layered social structures do emerge even when the ratio of social/work time does not favour high frequency social interaction. This suggests that quality rather than quantity of social interaction may be more critical, and that the social preference strategies, i.e. favour-the-few, are the key to emergence of stratified social networks. Hypothesis 3 (wellbeing from few intimate relationships) is supported by the high weighting of the wellbeing for the SBH pattern; however, wellbeing may also be promoted by many intimate relationships or a large number of less intimate relationships since high WB weighting promoted the many strong-ties and medium+ weak-ties patterns. High

alliance weighting with high wellbeing and low stress support hypothesis 4, but this was observed for many strong ties as well as the SBH pattern. Finally, hypothesis 5 (weak ties) is supported by the high resources weighting favouring the many weak-ties and medium+weak-ties patterns. Resource fitness was driven by the favour-the-many strategy.

Our simulation supports the observations that technology-mediated social interaction supplements but does not change the nature of human relationship, with the main function of social media being to maintain existing real-world social relationships [Ellison et al. 2007; Lampe et al. 2006]. The relative insensitivity of the model to more frequent interaction (cycles) suggests that more frequent social exchanges will not produce more strong trust relationships. However, social relationship structure was sensitive to population, so increased interaction frequencies with more alters, potentially afforded by social media, could increase the number of stronger trust relationships. Possible reasons why this conjecture is not supported by empirical evidence are that social disinhibition is increased by technology [Postmes et al. 2000], shown by the defect sensitivity in our model, or that the social media exchanges provide only minor trust increments which fail to overcome relationship waning rates. However, social media do appear to afford opportunities for developing and maintaining more weak ties, where interaction frequency and quality may be less important [Ellison et al. 2011; Boyd and Ellison 2007].

The emergence of stratified social structures, conforming to SBH, is influenced by both the selection rate and population size. In behaviour time, social learning of favour-the-few strategies would have to be rapid, possibly indicated by development of strong trust relationships in childhood [Hays 1989]. Increased social rewards allows SBH to appear in larger populations or with higher rewards and higher waning rates simulating the lower frequency of physical contact. Social media may facilitate more opportunities for social exchange; however, it is debatable whether the many technology-mediated interactions are actual exchanges rather than monitoring activities, e.g. wall posts/status update in Facebook [Lampe et al. 2006, 2007; Ellison and Boyd 2013].

Within the assumptions of the social trust model, selection favours two patterns in a wide variety of environmental contexts as defined by the fitness/reward space. The most frequent weak-ties pattern is driven by a mixed population of agents with medium- and weak-tie social preferences. The second medium-/weak-ties pattern is produced by the persistence of strong ties and staged strategy agents, with the social tendency to favour-the-few being a common competitive strategy appearing in nearly half (45%) of the simulations. Both of these patterns showed strong selection to increase work at the expense of social interaction, with the WS ratios migrating towards 1:1, so there appear to be two conflicting trends: one to promote stronger ties by favouring-the-few in social interactions, while the other increases work at the expense of social interaction. Increasing work favours weak-ties social structures without layers. This polarisation reflects the competing rewards that arise from maximising individual rewards by devoting more time to work for resources versus those accruing via social benefits. This echoes the work-life balance debate in society.

An interesting side effect of social preference selection is the reduction in defect rates. This suggests that non-cooperative behaviour is not favoured even when our trust model asserts that strong trust relationships are buffered against adverse interactions. Our interpretation is that non-cooperative behaviour is still perceived as a threat to relationship stability, hence it is selected against; and this will apply equally to Internet-mediated as well as face-to-face exchanges. Indeed, vigilance **on** social media to counter non-cooperative behaviour appears to be increasing [Ellison and Boyd 2013; Ellison et al. 2007].

The SBH-compliant and the many-/strong-ties patterns appear to be rare, appearing in only a small fraction of the settings (8%) in the overall fitness space, suggesting that they evolve only under very specific conditions, namely when social rewards are high (notably in terms of wellbeing and alliances) and social costs (risk and stress) are low. However, another explanation is competition from the WS balance, which was apparent in the more common medium-/weak-ties pattern. There appears to be strong selection pressure to favour less social and more work for resource unless fitness via wellbeing and alliances is prioritised. This indicates that structured sociality may only emerge when the pressure on social interaction time is less intense, and the impact of more work-resource gathering time for survival is less critical.

The clustering of fitness criteria observed for the SBH simulation runs supports our hypothesis that relationships at different levels of intimacy could emerge through different cost-benefit trade-offs. The two clusters that emerged had high alliance with relatively low wellbeing, balanced against high stress in one cluster and a mix of stress and risk in a second. This result agrees with the sympathy group trade-off. The other two SBH clusters map to the inner (support) group trade-off, with higher wellbeing and alliances balanced against lower risk and stress. Since relationships in the support group may also be expected to contribute to alliances, a high weighting on the alliance criterion is expected. The same clusters emerged in the many-/strong-ties pattern, but with a lower setting for risk and stress.

While the social trust model has provided a reasonably accurate account for the development of social relationships in behaviour and evolutionary time, there are many future directions for its application. The quality of interaction via social media could be modelled more explicitly to account for broadcast communication in contrast to dyadic exchanges. Further experiments with different mixes of face-to-face and computer-mediated-communication interactions could explore the role of social media on relationship formation and maintenance. The conflict between weak-tie rewards of information and networking resources versus socio-emotional benefits deserves further study. Our model contrasts with trust inference models which rely on users' ratings of alters [Golbeck and Hendler 2006] where trustworthiness can be assessed by a combination of ratings; in contrast, our social trust model takes a longer-term view where trust accumulates from social interactions and individual preferences in favouring particular alters to interact with. In conclusion, we have demonstrated the first trust model that provides a faithful account and explanation for the ontology of human social relationships from a psychological and evolutionary perspective. In the future we will apply the model for further simulations in social computing as well as social networks.

ACKNOWLEDGEMENTS

The authors wish to acknowledge support from the ESRC/EPSRC cognitive foresight programme project TESS, Developing Theory for Evolving Socio-technical Systems. RD's research is also supported by the British Academy centenary project (Lucy to Language) and the EUFP7 Socialnets project.

REFERENCES

- Baumeister, R.F. and Leary, M.R. 1995. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497-529.
- Binder, J., Roberts, S. and Sutcliffe, A.G. 2012. Closeness, loneliness, support: Core ties and significant ties in personal communities. *Social Networks*, 34(2), 206-214.
- Bonabeau, E., Dorigo, M. and Theraulaz, G. 1999 *Swarm intelligence: From natural to artificial systems*. Santa Fe: Santa Fe Institute. <http://www.swarm.org/wiki/main_page>
- Bowlby, J. 1990. *A secure base: Parent-child attachment and healthy human development*. London: Basic Books.

- Boyd, D. and Ellison, N.B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Briggs, P., Burford, B., DeAngeli, A. and Lynch, P. (2002). Trust in online advice. *Social Science Computer Review*, 20(3), 321-332.
- Brown, S.L. and Brown, R.M. 2009. Selective investment theory: Recasting the functional significance of close relationships. *Psychological Inquiry*, 17, 1-29.
- Buys, C.J. and Larson, K.L. 2008. Human sympathy groups. *Psychology Reports*, 45, 547-553.
- Cannon, J., Doney, P. and Mullen, M. 1998. National culture and the development of trust: The need for more data and more theory. *Academy of Management Review*, 24(1), 8-11.
- Castelfranchi, C. and Falcone, R. 2010. *Trust theory: A socio-cognitive and computational model*. Chichester: John Wiley.
- Crockford, C., Wittig, R.M. et al. 2008. Social stressors and coping mechanisms in wild female baboons (*Papio hamadryas ursinus*). *Hormones and Behavior*, 53, 254-265.
- Dawkins, R. 1976. *The selfish gene*. London: Oxford University Press.
- Dunbar, R.I.M. 1992. Coevolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16, 681-735.
- Dunbar, R.I.M. 1998. The social brain hypothesis. *Evolutionary Anthropology*, 6, 178-190.
- Dunbar, R.I.M. 2010. Brain and behaviour in primate evolution. In P.H. Kappeler and J. Silk (Eds), *Mind the gap: Tracing the origins of human universals*, pp. 315-330. Berlin: Springer.
- Dunbar, R.I.M. and Spoors, M. 1995. Social networks, support cliques, and kinship. *Human Nature*, 6, 273-290.
- Dunbar, R.I.M. and Stiller, J. 2007. Perspective-taking and social network size in humans. *Social Networks*, 29, 93-104.
- Dunbar, R.I.M., Korstjens, A. and Lehmann, J. 2009. Time as an ecological constraint. *Biological Reviews*, 84, 413-429.
- Ellison, N.B. and Boyd, D. (2013). Sociality through social network sites. In W.H. Dutton. (Ed.), *The Oxford handbook of Internet studies*. pp. 151-172. Oxford: Oxford University Press.
- Ellison, N.B., Steinfield, C. and Lampe, C. 2007. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12, 1143-1168.
- Ellison, N.B., Steinfield, C. and Lampe, C. 2011. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media and Society*, 13(6), 873-892.
- Falcone, R. and Castelfranchi, C. 2001a. Social trust: A cognitive approach. In C. Castelfranchi, and Y. Tan (Eds.), *Trust and deception in virtual societies*, pp. 55-90. Boston MA: Kluwer Academic Publishers.
- Falcone, R. and Castelfranchi, C. 2001b. The socio-cognitive dynamics of trust: Does trust create trust? In R. Falcone et al. (Eds.), *Trust in cybersocieties: Integrating the human and artificial perspectives*, pp. 55-72. Berlin: Springer.
- Fogg, B.J., Marshall, J. et al. 2001. What makes web sites credible? A report on a large quantitative study. In *CHI 2001 Conference Proceedings: Conference on Human Factors in Computing Systems, Seattle*. New York: ACM Press.
- Fowler, J.J. and Christakis, N. 2008. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337, a2338.
- Gilbert, N. and Troitzsch, K.G. 2005. *Simulation for the social scientist*. Milton Keynes: Open University Press.
- Golbeck, J. and Hendler, J. 2006. Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, 6(4), 497-529.
- Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology*, 78, 1360-1380.
- Hardy, C.L. and Van Vugt, M. 2006. Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32, 1402-1413.
- Hays, R.B. 1985. A longitudinal study of friendship development. *Journal of Personality and Social Psychology*, 48(4), 909-924.
- Hays, R.B. 1989. The day-to-day functioning of close versus casual friendships. *Journal of Social and Personal Relationships*, 6, 21-37.
- Heider, F. 1946. Attitudes and cognitive organization. *Journal of Psychology*, 21, 107-112.
- Hill, K., Hurtado, A.M. and Walker, R.S. 2007. High adult mortality among Hiwi hunter-gatherers: Implications for human evolution. *Journal of Human Evolution*, 52, 443-454.
- Hinde, R.A. and Stevenson-Hinde, J. 1990. Attachment: Biological, cultural and individual desiderata. *Human Development*, 33, 62-72.
- Hruschka, D.J. and Henrich, J. 2006. Friendship, cliquishness, and the emergence of cooperation. *Journal of Theoretical Biology*, 239, 1-15.
- Janssen M. 2006. Evolution of cooperation when feedback to reputation scores is voluntary. *Journal of Artificial Societies and Social Simulation*, 9(1) <<http://jasss.soc.surrey.ac.uk/9/1/17.html>>
- Joinson, A.N. 2008. Looking at, looking up or keeping up with people? *CHI 2008 Conference Proceedings: Conference on Human Factors in Computing Systems*, pp. 1027-1036. New York: ACM Press.
- Kramer, R.M. 1999. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, 50, 556-557.

- Lampe, C., Ellison, N. and Steinfield, C. 2006. A Face(book) in the crowd: Social searching vs. social browsing. *Proceedings: CSCW-2006*, pp. 167-170. New York: ACM Press.
- Lampe, C., Ellison, N. and Steinfield, C. 2007. A familiar Face(book): Profile elements as signals in an online social network. *Proceedings: 2007 Conference on Human Factors in Computing Systems, CHI 2007, San Jose*. New York: ACM Press.
- Larzelere, R.J. and Huston, T.L. 1980. The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, 42(August), 595-604.
- Lu, Y., Roberts, S. et al. 2009. Size matters: Variation in personal network size, personality and effect on information transmission. *Proceedings, 2009 International Conference on Computational Science and Engineering*, vol. 4, pp. 188-193. Los Alamitos: IEEE Computer Society Press.
- McCabe, K.A. (2002). A cognitive theory of reciprocal exchange. In E. Ostrom and J. Walker (Eds), *Trust and reciprocity: Interdisciplinary lessons from experimental research*, pp. 147-169. New York: Russel Sage Foundation.
- Mitchell, P. 1997. *Introduction to theory of mind: Children, autism and apes*. London: Arnold.
- Nowak, M. and Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature*, 383, 537-577.
- Nowak, M. and Sigmund, K. 2005. Evolution of indirect reciprocity, *Nature*, 437, 1291-1298.
- Ostrom, E. 2002. Towards a behavioural theory linking trust, reciprocity and reputation. In E. Ostrom and J. Walker (Eds), *Trust and reciprocity: Interdisciplinary lessons from experimental research*, pp. 19-79. New York: Russel Sage Foundation.
- Oswald, D.L., Clark, E.M. and Kelly, C.M. 2004. Friendship maintenance: An analysis of individual and dyad behaviours. *Journal of Social and Clinical Psychology*, 23(3), 413-441.
- Postmes, T., Spears, R. and Lea, M. 2000. The formation of group norms in computer-mediated communication. *Human Communication Research*, 26(3), 341-371.
- Rempel, J.K., Holmes, J.G. and Zanna, M.P. 1985. Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95-112.
- Riegelsberger, J., Sasse, M.A. and McCarthy, J.D. 2003. Shiny happy people building trust? Photos on e-commerce websites and consumer trust. In *CHI 2003 Conference Proceedings: Conference on Human Factors in Computing Systems*, pp. 121-128. New York: ACM Press.
- Roberts, S.G.B. and Dunbar, R.I.M. 2011a. Communication in social networks: Effects of kinship, network size and emotional closeness. *Personal Relationships*, 18, 439-452.
- Roberts, S.G.B. and Dunbar, R.I.M. 2011b. The costs of family and friends: An 18-month longitudinal study of relationship maintenance and decay. *Evolution and Human Behavior*, 32, 186-197.
- Roberts, S.B.G. and Dunbar, R.I.M. 2012. Communication in social networks: Effects of kinship, network size and emotional closeness. *Personal Relationships*, 18, 439-452.
- Roberts, S.G.B. and Renwick, J.S. 2003. The development of cooperative relationships: An experiment. *Proceedings of the Royal Society of London*, 270, 2279-2283.
- Rose, S. and Serafica, F. 1986. Keeping and ending casual, close and best friendships. *Journal of Social and Personal Relationships*, 3(3), 275-288.
- Rotenberg, K.J. and Morgan, C.J. 1995. Development of a scale to measure individual differences in children's trust-value basis of friendship. *Journal of Genetic Psychology*, 156(4), 489-502.
- Rotter, J.B. 1971. Generalised expectancies for interpersonal trust. *American Psychologist*, 26, 443-452.
- Sutcliffe, A.G. 2006. Trust: From cognition to conceptual models and design. In *Proceedings of Advanced Information Systems Engineering 18th International Conference, CAiSE 2006*. Berlin: Springer.
- Sutcliffe, A.G. and Wang D. 2012. Computational modelling of trust and social relationships. *Journal of Artificial Societies and Social Simulation*, 15(1), 3.
- Sutcliffe, A.G., Dunbar, R.I.M., Binder, J. and Arrow, H. 2012. Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology*, 103(2), 149-68.
- Sutcliffe, A.G., Gonzalez, V., Binder J. and Nevarez, G. 2011. Social mediating technologies: Social affordances and functionalities. *International Journal of Human Computer Interaction*, 27(11), 1036-1065.
- Wellman, B., Hogan, B. et al. 2006. Connected lives: The project. In P. Purcell (Ed.) *Networked Neighborhoods*. Berlin: Springer.
- Whitmeyer J.M. 2002. A deductive approach to friendship networks. *Journal of Mathematical Sociology*, 26(3), 147-165.