

The Value of Trust

Robert Ghanea-Hercock
Said Business School, Oxford,
robert.ghanea-hercock@bt.com

Abstract

What value should a corporation place on trust within its internal structure? And to what extent does the perceived trustworthiness of the corporation count in regard to its external interactions? It is proposed that these questions are in fact pivotal to the longevity and innovation capacity of any organization, whether commercial or public. In this paper we utilize a multi-agent simulation to explore the dynamics of trust interactions, and the impact on the cohesion of the resulting agent society. These results are then discussed in the light of the questions posed above, i.e. what correlation exists between an organizations efficiency and survival, and the degree of trust in the overall system. In this context, we are addressing trust both between individual actors and with external groups or agents who may interact with the organization.

In order to study the impact of trust, software agents provide a useful means to analyze the formation of trust within groups. However, the theory of trust dynamics remains a poorly understood domain, [see Nowak & May 1992]. In particular, this work considers how the cost of trust formation between agents shapes the cohesion and efficiency of the agent society.

The simulation results indicate that if an organization successfully creates a high level of inter-agent trust then it possesses a significant degree of systemic robustness. However, if the internal trust level falls below a critical threshold then the cohesion of the collective group becomes highly fragile. The conclusion is that organizations that fail to sustain and cultivate an internal culture of trust are at risk of a collapse of innovation and efficiency, which ultimately leads to organizational failure.

Keywords: Corporations, Trust, & Innovation

1. Introduction

The work by Fukuyama [1995] cast light on a critical aspect of organizational theory that raised the issue of inter-agent trust as a vital component of social and economic stability, and more importantly efficiency. Similarly, the related text on this topic by Collins and Porras [2005] elucidated the ‘social chemistry’ that enabled some corporations to survive for extended periods of time. For example, in studying the pharmaceutical group Merck & Company, they discuss how:

*“Merck, in fact, epitomizes the ideological nature – the pragmatic idealism – of highly visionary companies. Our research showed that a fundamental element in the “ticking clock” of a visionary company is a **core ideology** – core values and sense of purpose beyond just making money – that guides and inspires people throughout the organization and remains relatively fixed for long periods of time.”*

[Collins & Porras, 2005, p.48]

In this paper we argue that the root of these ‘values and purpose’ is the degree of social capital that exists within an organization. Or as Putnam states, it is the: *“connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them.”*

[*Bowling Alone*, Putnam, 2001]

The premise is that companies with strong internal cohesion and rich social capital are much more robust in their ability to assimilate and drive innovation. Greater cohesion effectively provides an organization with ‘shock absorbers’, such that it can adapt and flex around disruptive innovations. In addition if the employees of an organization feel trusted and operate in a high trust environment then open innovation and the genuine exchange of information is possible. Although, building and sustaining corporate social cohesion is an immense challenge. However, changes to commercial norms of behavior

can occur. For example, in a state that Fukuyama assessed as having relatively weak prior social capital, i.e. China, the foreign corporations that began operating there since the 1990's, have in effect seeded a culture of large-scale enterprise management, that has now been assimilated by the indigenous companies. As the basis of social capital is the trust networks that exist between agents, we need to understand what enables increased trust and reciprocity within an organization. The tool selected for this is a computational multi-agent simulation which is presented in the following section.

1.1 Agent-based Models

Multi-agent simulations (MAS) have been used to analyze a wide range of economic and social problems. The value of this approach is well articulated in a paper by Epstein [Epstein 2008]. (See the web site <http://jasss.soc.surrey.ac.uk/JASSS.html> for a range of current articles on social, economic, and organizational MAS applications.) While a number of mechanisms have been utilised to develop agent reputation systems [Abdul-Rahman & Hailes 2000], and [Yu & Singh 2002], there is still a lack of understanding in terms of how such processes evolve over time, or the dynamic behaviour of such systems. For example, the reputation development between agent brokers in e-commerce systems [Braynov & Sandholm 1999].

1.2 Trust

This section briefly outlines what is meant by trust in the context of software agents. A significant body of work has addressed the notion of trust in computational systems, such as [Castelfranchi & Falcone 1998], or the early work on evidence based reasoning [Shafer 1976]. A useful working definition of trust has been suggested by Marsh:

"...trust, (or symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before he can monitor such action (or independently of his capacity to monitor it) and in a context in which it affects his own action." [Marsh 1999]

Of specific interest to this work is the study of agent group formation managed by processes of trust, such as that identified by [Griffiths & Luck 2003], who suggest a clan style agent collective process, within which agents experience mutual benefit. A second important approach is that of [Jonker & Treur 1999], which considers the dynamic evolution and adaptation of trust within individual agents. The focus of this work is similarly concerned with how an agent modifies its own trust perception of social interactions and events. For example:

*"Each event that can influence the degree of trust is interpreted by the agent to be either a **trust-negative** experience or a **trust-positive** experience. If the event is interpreted to be a trust-negative experience the agent will lose trust to some degree, if it is interpreted to be a trust-positive, the agent will gain trust to some degree"*.

[Jonker & Treur 1999]

This adaptive shift in an agent's degree of trust forms the basis for the trust algorithm presented in section 3 of this paper.

1.3 Hypothesis: Passive Trust

In many classes of agent populations the individual agents do not possess the time or resources to perform a complete assessment of the trustworthiness of every agent they interact with. Such an assessment can be defined as an "*active trust*" process. In contrast, it is proposed that it is economically more efficient for most agents to utilise a "*passive trust*" process. By this is meant that an agent preferentially selects which agents to interact with on the basis of a similarity metric applied to the other agent. (For comparison, work on *homophily* between economic agents, illustrates how the reason for group formation and parochialism, is precisely the lower cost of transaction that occurs within a group [Bowles & Gintis 1998]). In contrast if an agent has control of a centralized or global resource, then it is in principle possible to perform an active trust assessment of every agent that requests an interaction. This raises the issue of exactly how an agent should apply a suitable payoff-matrix to determine when to apply a more active trust

procedure. This hypothesis of passive trust has been tested via a multi-agent simulation, the results of which are described in the remainder of the paper.

Section 2 describes the background theory to trust formation in agent systems. Section 3 covers the specific model developed for this work. Sections 4 and 5 outline the results and conclusion from this model.

2. The Cost of Trust

The question of interest is therefore how much resource should an individual agent assign to determine the trustworthiness of another agent. In particular, are there any procedures available which will minimize this cost to the agent? This section looks at the two major alternatives available to an agent.

Active Trust

The process of determining whether an agent is trustworthy will always incur a cost to the agent making the evaluation. For example either the agent must consult a central or distributed database that lists agents trust rating, or it must perform some internal assessment. Any internal assessment normally requires a long term memory of previous interactions with the agent and/or information from N other agents that may have referral information regarding the trust status of the agent in question, [Yu & Singh 2002]. In the case of a software agent this translates to the consumption of machine cycles, working memory, data storage, and communication bandwidth/access to make the requests. This point is reinforced by [Ramchurn et al 2003], where a sophisticated soft inferencing mechanism is used by agents to evaluate the trustworthiness of other agents prior to entering a contractual state. They also point out the cost issues associated with an agent making such a judgment. The implication is that trust operates on a macro-scale in determining the strength of an interaction process between agents.

In human terms the cost of actively validating the trustworthiness of every transaction or interaction is frequently unsustainable. Hence, human actors rely heavily on the second form of trust inferencing, i.e. Passive Trust.

Passive Trust

In contrast, if an agent is applying a passive trust policy, then all it needs is to request, or sense, a unique identifying tag from another agent, in order to grant it some degree of trust. (Where a tag is defined as any unique attribute displayed by an agent; see Holland 1993, or Riolo 1997 for a broader discussion of tag processes in agent systems.)

This process is frequently utilised within human economic transactions, either via the simple mechanism of badges on individuals, or logos on corporate products. Indeed such logos are also fiercely protected by legal sanctions to maintain their reputation value. For example, a human buyer interested in purchasing a product may take two approaches, either to perform an extensive study of the past history and reliability of the product, or to simply check the logo and compare it with a publicly known list of popular vendors. In the second case they are utilizing passive trust in assuming that the logo still correlates with a quality product. Of course this process also undergoes dynamic evolution, as logos can gain and lose reputation and value. As in the recent case of Toyota and sticking accelerator pedals.

The key aspect of this process relates to organizational efficiency. If the social norms and operational processes reinforce trust within a group, then it greatly reduces the cost of inter-agent transactions. The same principle applies in relation to external transactions. Hence, organizational efficiency is heavily dependent on the trust dynamics operating within the system.

3. Dynamic Evolution of Trust

This section outlines a simple MAS model that explores the impact of a passive trust mechanism on group cohesion. An agent's interaction in any complex environment requires a continuous reassessment of the degree of trust it should assign to external agents and events. Any policy based on a fixed degree of trust is liable to incur serious costs for an agent. The underlying assumptions made in this simulation are:

- a) Trust is assigned as a continuous variable internal to each agent.
- b) Positive events result in an agent increasing its degree of trust.
- c) Negative events result in an agent decreasing its degree of trust.
- d) Agents apply a threshold function to determine whether to trust another agent.
- e) The threshold value is a function of the agents' current degree of trust.

As suggested in the paper by [Griffiths and Luck 2003] agents should offer preferential trust to agents within a specific social clan or group with which they are affiliated. It is proposed that by selectively trading within restricted social groups an agent can save on the expenditure of resources required to determine whether to interact with another agent; compared to the process of *active trust* assessment. This process takes advantage of the contextual situation created by a trustworthy social group.

3.1 Experimental Model

This section describes the simulation model used to test the hypothesis of the paper. Using the REPAST agent simulation platform [<http://repast.sourceforge.net>], a population of agents was created with the following attributes:

- Vision – range of local cells the agent can perceive and directly interact with.
- Metabolism – agents consume energy at a rate specified by the metabolism.
- Trust parameter – defined as the degree to which an agent will trust agents within its vision.
- Threshold Function – a step function used by the agent to decide whether to interact or not.
- Group Identity parameter – an initially random integer in the range 0-100 that defines an arbitrary identity tag.
- Energy parameter – the value of an agent's current energy parameter.
- Mailbox of received messages – an internal list of recently received messages to allow asynchronous exchange between the agents.

The use of an asynchronous messaging process allows for more flexible interaction patterns, such as agents moving in the environment or irregular spatial topologies. The simulation selects a number of agents at random each time step and calls the execution method of each agent. The rules applied by an agent are:

- i) Broadcast a message to all agents within the neighbourhood defined by its vision.
- ii) Parse all messages from the mailbox and respond; this is currently limited to reading the group ID tag of the sending agent.
- iii) If the ID tag of the sending agent is within the social threshold specified by the agent's internal trust parameter then invoke a positive trust response.
- iv) If the ID tag is outside the agent's threshold range then invoke a negative trust response.
- v) If this agents Energy has fallen to 0, then reset the agent to its initial state, (usually a random trust value with a new randomized tag).

If an agent experiences a positive trust response event the agent gains one unit of energy and increments its trust value by one. As a result it becomes more trusting and is able to interact with a higher percentage of the neighbouring agents. A negative trust response involves the agent losing one unit of energy and decrementing its trust value by one. Hence the agent has incurred a negative cost as a result of interacting with an agent outside its trust domain. It therefore becomes less trusting, which reduces the percentage of agents it may interact with in future.

Clearly the iteration of this process can lead to cycles of positive or negative feedback for each agent, which leads to either a global low or high trust regime. This final state depends strongly on the initial conditions, as discussed in the following results section. (It would also be interesting to use a non-uniform network topology to determine whether the topology of the process has a major effect. Related work on trust dynamics by [Yu & Singh 2003] has considered the evolution of trust on small-world type networks.)

4. Results

A number of experiments were performed using the specified model in the REPAST agent simulation tool. A population of 400 agents was initialized in a regular grid, with the parameters defined in section 3. The vision and metabolism parameters were both set to 1, while the tag parameter is initialized as a random integer between 0-100. The simulation colour-codes each agent according to the tag value and this is mapped onto an integer colour scale, for visualisation purposes, (see figure 1).

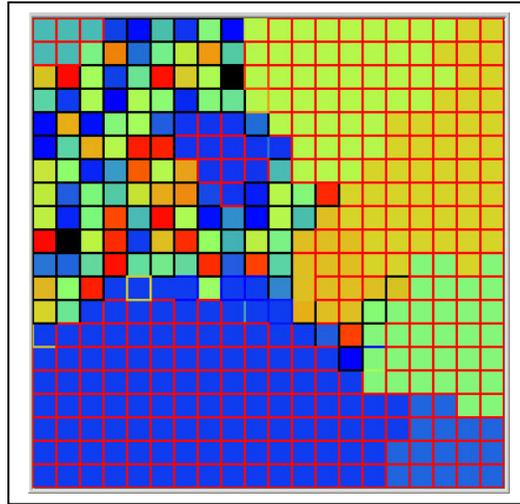


Figure 1 Screenshot from the REPAST simulation, showing a set of 2 stable agent groups. The lower left quadrant area shows a group of agents in a low-trust state.

A range of experiments were performed using different initial values for the trust parameter. In addition, alternative update strategies for shifting the trust value were tested and whether it benefits the agent to actively change its tag parameter to match that of neighbouring agents. (In related work we have examined the issue of agents defecting by intentionally displaying tags that maximize their trust rating [Ghanea-Hercock 2007].)

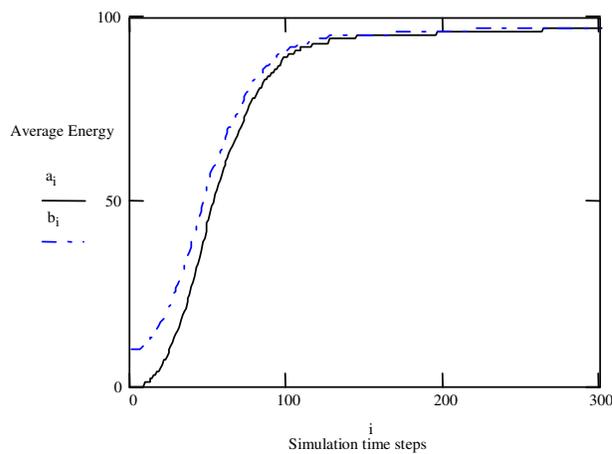


Figure 2 A typical response of the average energy and trust for a population of 400 agents after 1000 time steps. (Initial trust threshold value = 10). (a_i = average energy, b_i = average trust).

The evolution of the agents state is illustrated in figure 2, which shows the average energy and trust response over 300 time steps. In this example, the majority of the agent population has converged on a high trust state.

A second set of experiments investigated the stability of the social groups once formed. In this instance via an external suppression of the agents trust parameter (see fig.3). At some time-step after the emergence of a stable set of groups the trust parameter of all agents was reduced by some percentage.

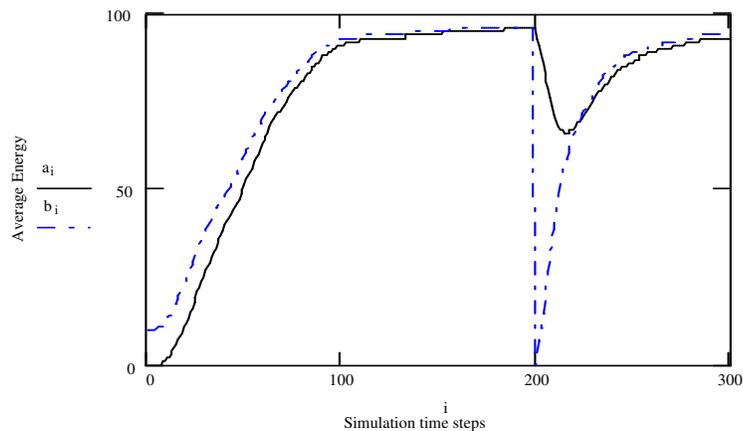


Figure 3. Adaptive response of agent population to global perturbation, i.e. all of the agents trust values are set = 0 for 2 time steps, at time step 200. (a_i = average energy – solid line, b_i = average trust – dashed line).

In this example (figure 3) the agent population rapidly recovers to a high trust and high average energy state. A final set of experiments were performed to determine the impact of the trust parameter on the convergence of the agent population to a stable trusting group. (This was measured by the average agent energy value, as the system may converge on a variable number of agent groups.) The key metrics for the experiments were the average trust and energy values for all of the agents; (as measured after 1000 simulation time steps.) These were used to indicate whether the agents were efficiently interacting and whether the population had converged to a low or high trust regime.

With an initial trust threshold value greater than 5 the agent population rapidly converges to a set of stable high trust groups. For initial trust threshold values lower than $T=5$, the system still converges to a stable set of groups, but the time required increases by more than an order of magnitude, i.e. an average time scale greater than the normal 4000 simulation time steps.

Based on these experiments the following set of observations was made:

a) If the initial average trust level of the agent population is below a critical trust threshold value then the population will remain in a low trust state, as indicated in figure 4. With correspondingly low market efficiency, defined as the average population energy in this model.

b) Once a set of high-trust agent groups have formed the system as a whole exhibits a high degree of robustness under external perturbation of the trust parameter. Even if we externally suppress the trust of every agent for a short duration, the system adaptively recovers to a high trust state, (see figure 3). Hence the spatially correlated set of tags within a group acts as a collective memory and leads to a rapid reformation of a stable group. This implies that a strong corporate identity can confer a significant level of robustness in the face of system perturbation.

c) In contrast, if the trust level is suppressed before high-trust groups have stabilized then a perturbation can lead to collapse into a permanent low-trust configuration.

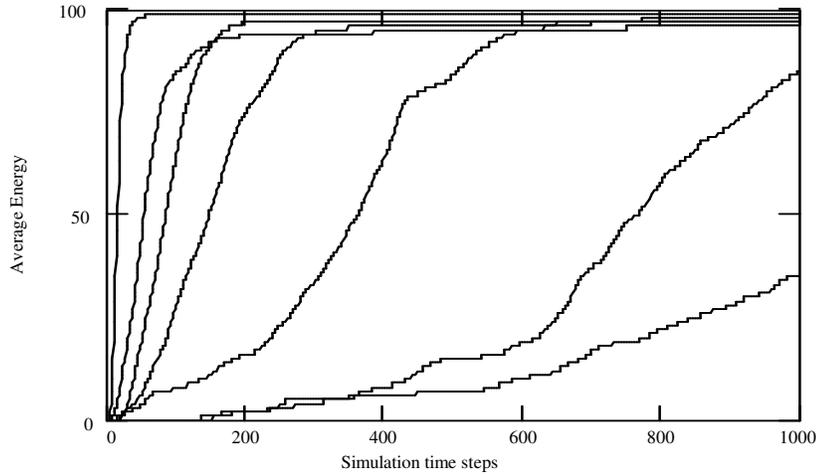


Figure 4 Plot of average energy as a function of decreasing initial trust threshold (from T=11 to T=5), illustrating a critical trust threshold between T=7 and T=6, (with a population of 400 agents). The lowest curve is for T=5.

4.1 Analysis

Based on the observed results a comparative dynamics model was explored that demonstrates closely related dynamical behaviour, i.e. the generalized Ising model [Ising 1925]. (For an introduction to the model see [Cipra 1987]). The Ising model imitates behaviour in which individual elements modify their behavior so as to conform to the behavior of other individuals in their vicinity. Originally used to explain empirically observed results for ferromagnetic materials, it has since been widely applied to multi-agent systems [Sznajd & Sznajd 2000], [Sherrington et al 2000] and economic models [Schultz 2003].

The model specifies that given a set of N individuals arranged in a lattice, each individual can be in one of two different states, for example, +1 and -1. Using the simplest form of the model with spin vector interaction restricted to nearest neighbours, an expression for the energy of any particular state is given by:

$$E = -J \sum_{\langle i,j \rangle} S_i S_j - B \sum_k S_k \quad \text{Eq.1}$$

Where: $S(j)$ (written with a subscript "j" in the equation) is the value of the spin at the j_{th} site in the lattice, with $S = +1$ if the spin is pointing Up and $S = -1$ if the spin is pointing Down. The $\langle i,j \rangle$ subscript on the summation symbol indicates that the spin-spin interaction term, $S(i)*S(j)$ is added up over all possible nearest neighbour pairs.

The constant J has dimensions of energy and it measures the strength of the spin-spin interaction. If J is positive the energy is lowered when adjacent spins are aligned. The constant B (again, an energy), indicates an additional interaction of the individual spins with some external magnetic field.

We can then compare the behaviour of the agent population, as shown in figure 2, with the dynamics of the Ising model. This is a potentially useful feature of this comparative model, as it offers a precise definition of trust between software agents, as defined by the Ising equation for interaction (Eq.1).

The metabolic parameter in the agent model can be equated with the constant B in the Ising model, i.e. an external interaction field that affects the agent's energy state. The resulting phase transition shown in figure 2 is a distinct characteristic of the Ising model and related neighbourhood interaction models. The agent model however is complicated by the large number of states per agent, and the ability of each agent to vary the strength of its interaction with its neighbours via the trust parameter T . (These effects become particularly manifest at low trust parameter values, and result in multiple inflection points in the phase transition curve). However, the general dynamics of the Ising model remains a useful potential tool for

analyzing the dynamics exhibited by this class of multi-agent interaction model, in a qualitative and quantitative manner.

4.2 Future Work

A number of areas for further work need to be developed using this model. For example, adding some form of trading interaction between the agents to simulate a model economic process is required. This would enable a more grounded cost and utility function based evaluation of the trust process. The key extension, however, will be to add an active trust assessment model, so that agents can select between the two strategies according to the situation, and ideally utilise a learning process to optimize the balance between them. In addition we have started work on modelling intentional defection by the agents and analyzing the results from a game-theoretic perspective, and a consideration of the pay-off matrix impact versus inter-agent trust evaluations. Of course this adds greatly to the complexity of the model and system dynamics, [Nowak & May 1992].

5. Conclusion

The widespread practice of parochial selection in human social groups appears to support the hypothesis of this paper, i.e. that agents are frequently unable or unwilling to support the cost of an active assessment of trust for every interaction they encounter. By using a simple low cost process of passive trust assessment based on group defined tags, a population of agents can achieve a stable and adaptive high-trust community, with minimal information processing or computing resources. This work aims to offer some insight into the formation of high and low trust groups within organizations. In particular why such groups tend to exhibit long term persistence, and why such strong parochialism occurs in the first place. It also suggests that if sufficient numbers of agents in a local area can be moved past a critical level of trust then a positive feedback mechanism can then shift the entire population into a high-trust state.

From the perspective of self-organization, the results also show how trust dynamics operate as self-sustaining feedback loops. This has serious consequences for both strategy and policy, as even small decisions that impact trust will be magnified rapidly across the organization.

Second, if an organizations internal trust levels are low then any innovation is slow to propagate through the management structure. From an external perspective other groups are also reluctant to share or develop cooperative innovation processes, where the trust status of a potential partner is perceived as being weak.

Third, companies generally accept that brand is important, but if we consider it in the light of a passive trust model, then brand perception becomes vital to projecting trust. Within an organization the role of a distinct corporate identity is also vital, in enabling smooth and efficient trust interactions between employees and other agents.

Finally, the long-term survival of any corporate, or organizational unit, is deeply influenced by the trust dynamics that operate within it. Policies which undermine group trust will have a crippling impact on the ability to generate and absorb innovation and hence the probability of survival.

Acknowledgements

This work was conducted via a Visiting Fellowship held at the Complex Agent-Based Dynamic Networks group in the Said Business School, Oxford.

6. References

1. Abdul-Rahman A., and Hailes S. "Supporting Trust in Virtual Communities". In: Hawaii Int. Conference on System Sciences 33, Maui, Hawaii, Jan. 2000.
2. Bowles S., and Gintis H., "Optimal parochialism: The dynamics of trust and exclusion in networks". Working paper, University of Massachusetts, Amherst, 1998.
3. Braynov S., and Sandholm T., "Contracting with Uncertain Level of Trust", Proceedings of the first ACM conference on Electronic commerce, November 3-5, 1999.
4. Castelfranchi C., and Falcone R., "Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98), pages 72–79, Paris, 1998.
5. Cipra, B. A. "An Introduction to the Ising Model." Amer. Math. Monthly **94**, 937-959, 1987.
6. Collins, James, C., and Porras, Jerry, I., "Built to Last: Successful Habits of Visionary Companies", (Hardcover), Random House Business Books, New edition, 2005.
7. Epstein, Joshua, M., "Why Model?", Journal of Artificial Societies and Social Simulation, vol.11 no.4 <http://jasss.soc.surrey.ac.uk/11/4/12.html>, 2008.
8. Fukuyama F., "Trust: The Social Virtues and the Creation of Prosperity". New York: FP, 1995.
9. Ghanea-Hercock R.A., "Phobos: An Agent-based User Authentication System", IEEE Intelligent Systems, **18**(3): 67-73 (2003), June 2003.
10. Ghanea-Hercock R.A., "Survival in Cyberspace", Information Security Technical Report, – Elsevier pub. 2007.
11. Griffiths N., and Luck M., "Coalition Formation through Motivation and Trust", International Conference of Autonomous Agents and Multi-Agent Systems, Melbourne, Australia, 2003.
12. Holland J., "The Effects of Labels (Tags) on Social Interactions", Santa Fe Institute Working papers 93-10-064, Santa Fe, NM, 1993.
13. Ising, E. "Beitrag zur Theorie des Ferromagnetismus." Zeitschr.. f. Physik **31**, 253-258, 1925. [German].
14. Jonker C., and Treur J., "Formal analysis of models for the dynamics of trust based on experiences". In Multi-Agent System Engineering, Proceedings of the 9th European Workshop on Modeling Autonomous Agents in a Multi-Agent World, MAAMAW'99. LNAI 1647, 1999.
15. Maes P., Guttman R. and Moukas A., "Agents that Buy and Sell: Transforming Commerce as we Know It." Communications of the ACM, Mar. 1999 Issue.
16. Marsh, S., "Formalising Trust as a Computational Concept". Ph.D. Thesis. Department of Mathematics and Computer Science, University of Stirling, <http://citeseer.nj.nec.com/marsh94formalising.html>, 1999.
17. Nowak, M.A., and May, R.M., "Evolutionary Games and Spatial Chaos", Nature, 359(6398), 29th October, pp. 826-829, 1992.
18. Putnam, Robert, "Bowling Alone: The Collapse and Revival of American Community" (Paperback), Simon & Schuster Ltd; New edition, 2001.
19. Ramchurn S., Jennings N., Sierra C., and Godo L., "A Computational Trust Model for Multi-Agent Interactions based on Confidence and Reputation", International Conference of Autonomous Agents and Multi-Agent Systems, Melbourne, Australia, 2003.
20. Riolo R., "The Effects and Evolution of Tag-Mediated Selection of Partners in Populations Playing the Iterated Prisoner's Dilemma", Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97), pub. Morgan Kaufmann, San Francisco, CA, ed. Thomas B., 1997.
21. Schultz M., "Statistical Physics and Economics Concepts, Tools and Applications", Series: Springer Tracts in Modern Physics, Vol. 184 M. Germany, pub. 2003.
22. Shafer G. "A Mathematical Theory of Evidence". Princeton University Press, Princeton, NJ, 1976.

23. Sherrington D., Garrahan J. and Moro E., "Statistical physics of adaptive correlation of agents in a market", cond-mat/0010455, 2000.
24. Sznajd-Weron K., and Sznajd J., Int. J. Mod. Phys. C 11, 1157, 2000.
25. Yu B., and Singh M., "An Evidential Model of Distributed Reputation Management". International Conference of Autonomous Agents and Multi-Agent Systems, Bologna Italy, 2002.
26. Yu B., and Singh M.P., "Detecting Deception in Reputation Management", Proceedings of Second International Joint Conference on Autonomous Agents and Multi-Agent Systems, 2003.