Analysing the Relationship between Risk and Trust

Audun Jøsang[1] and Stéphane Lo Presti[2]

[1] DSTC * * *, Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001,
Australia.
`ajosang@dstc.edu.au`
[2] University of Southampton [†], School of Electronics and Computer Science,
Southampton SO17 1BJ, United Kingdom.
`splp@ecs.soton.ac.uk`

Abstract. Among the various human factors impinging upon making a decision in an uncertain environment, risk and trust are surely crucial ones. Several models for trust have been proposed in the literature but few explicitly take risk into account. This paper analyses the relationship between the two concepts by first looking at how a decision is made to enter into a transaction based on the risk information. We then draw a model of the invested fraction of the capital function of a decision surface. We finally define a model of trust composed of a *reliability trust* as the probability of transaction success and a *decision trust* derived from the decision surface.

1 Introduction

Manifestations of trust are easy to recognise because we experience and rely on it every day. At the same time it is quite challenging to define the term because it is being used with a variety of meanings and in many different contexts [12], what usually lead to confusion. For the purpose of this study the following working definition inspired by McKnight and Chervany's work [12] will be used:

Definition 1 (Trust). *Trust is the extent to which one party is willing to depend on somebody, or something, in a given situation with a feeling of relative security, even though negative consequences are possible.*

Although relatively general, this definition explicitly and implicitly includes the basic ingredients of trust. The term *situation* enables this definition to be adapted to most needs, and thus be general enough to be used in uncertain and changing environments. The definition acknowledges the subjective nature of trust by relying on one's *willingness* and *relative* security. The aspect of dependence is implicitly complemented by uncertainty through *possibility* and by risk through *negative consequences*.

Risk emerges for example when the value at stake in a transaction is high, or when this transaction has a critical role in the security or the safety of a system. It can be seen as the anticipated hazard following from a fault in or an attack of the system and can be measured by the consequences of this event.

The question regarding whether adequate models of risk and trust can be designed is still open at the present time. This ensues from the fact that these two notions encompass so many aspects of our life that their understanding is made difficult by the scale and the subjectivity of the task. Furthermore, these notions intrinsically rely on uncertainty and unpredictability, what complicates even more their modelling. Nevertheless, many models and approaches have been proposed to delimit, to reason and to solve a part of the problem that trust and risk constitute.

There are at the moment few trust systems and models that explicitly take the risk factor into account [8]. In most trust systems considering risk, the user must explicitly handle the relationship between risk and trust by combining the various ingredients that the system provides. At the same time, all those systems acknowledge the intuitive observation that the two notions are in an inverse relationship, i.e. low value transactions are associated to high risk and low trust levels and vice versa, or, similarly, risk and trust pull in opposite directions to determine a user's acceptance of a partner [13].

Falcone and Castelfranchi (2001) [6] recognise that having high trust in a person is not necessarily enough to decide to enter into a situation of dependence on that person. In [6] they write: *"For example it is possible that the value of the damage per se (in case of failure) is too high to choose a given decision branch, and this independently either from the probability of the failure (even if it is very low) or from the possible payoff (even if it is very high). In other words, that danger might seem to the agent an intolerable risk."*

Povey (1999) [14] introduces the concept of risk in McKnight and Chervany's work. Risk is exposed by the Trusting Behaviour and influences the Trusting Intentions and possibly the Situational Decision to Trust. Dimitrakos (2002) [5] presents this schema with a slightly different, and corrected, point of view. Trust metrics, costs and utility functions are introduced as parameters of an algorithm that produces the trust policy for a given trusting decision. Nevertheless, this work lacks a quantitative definition of the various involved measures and lacks examples of application of this generic algorithm.

The SECURE project [4] analyses a notion of trust that is *"inherently linked to risk"*. Risk is evaluated on every possible outcome of a particular action and is represented as a family of cost-PDFs (Probability Density Function) parameterized by the outcome's intrinsic cost. The considered action is then analysed by a trust engine to compute multidimensional trust information which is then used by a risk engine to select one cost-PDF. The decision to take the action is then made by applying a user-defined policy to select one of the possible outcomes' cost-PDFs.

The system described by Manchala (1998) [10] avoids expressing measures of trust directly, and instead develops a model based on trust-related variables such as the cost of the transaction and its history, and defines risk-trust decision matrices as illustrated in Figure 1. The risk-trust matrices are then used together with fuzzy logic inference rules to determine whether or not to transact with a particular party.
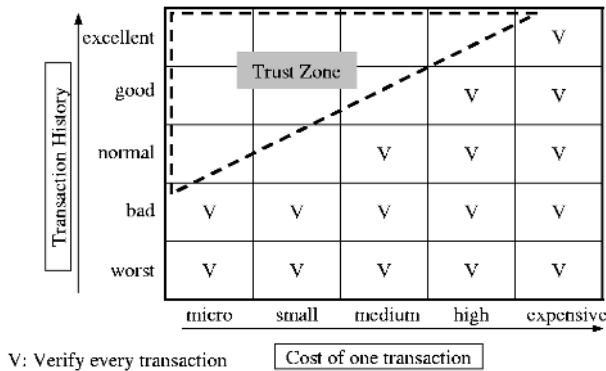
Fig. 1. Risk-trust matrix (from Manchala (1998) [10]).

In this paper, we expand on Manchala's model of trust with the intention of refining the relationship between trust and risk. Section 2 analyses how risk influences the decision-making process by calculating some of the factors impinging upon the outcome of the transaction. These factors are the expected gain and the fraction of the capital that is invested. Section 3 follows by deriving two trust factors from the previous elements: reliability and decision trust. In Section 4, we conclude by summarising our approach and discussing further work.

2   Decisions and Risk

Risk and trust are two tools for making decisions in an uncertain environment. Though central to many works, these two indicators only have semantics in the context of a decision that an agent is taking. An agent, here, can equivalently be a human being (e.g. a stockbroker) or a program (e.g. a software agent), whose owner (another agent) has delegated the decision-making process for a particular kind of interaction.

We focus on transactions rather than general interactions. This abstraction is not a limitation but rather a point of view on interactions, since most interactions can be modelled by transactions. Collaborations can be viewed as a group of transactions, one for each collaborator. The case of an attack by a malicious agent is a degenerated case where the transaction is abused by the attacker who invests fake income. Lastly, dependability can be considered as a combination of the various transactions between the agents so that the transactions' history and their overall effect are summarized.

Since risk is involved, we assume that numerical information is available from the transaction context to compute the level of risk. Practically these values may be hard to determine, since many factors of the transaction need to be taken into account [7,1], and financial modelling may not be suited to all the transaction contexts. For the sake of simplicity, we will limit ourselves to simple financial transaction contexts, but without loss of generality in our explorative approach.

In classical gambling theory the *expected monetary value EV* of a gamble with n mutually exclusive and exhaustive possible outcomes can be expressed as:

$$EV = I \sum_{i=1}^{n} p_i G_i \tag{1}$$

where p_i is the probability of outcome i and G_i is the gain factor on the monetary investment (or bet) I in case of outcome i.

However in many cases the utility is not the same as monetary value, and *expected utility EU* is introduced to express the personal preferences of the agent. In classical utility theory the expected utility can be expressed as a linear function of the probabilities:

$$EU = \sum_{i=1}^{n} p_i u(IG_i) \tag{2}$$

where u is an *a priori* non-linear function of monetary value. In traditional EU theory the shape of the utility function determines risk attitudes. For example, the agent would be risk averse if u is a concave function, meaning that, at a particular moment, utility gain from a certain transaction outcome is less that the actual monetary value of the same outcome. Considering that a utility function is identified only up to two constants (origin and units) [11], the concavity condition can be simplified to: $u(IG) < IG$ for risk aversion behaviour; and $u(IG) > IG$ for risk seeking behaviour.

However, studies (e.g. [2]) show that people tend to be risk seeking for small values of p, except if they face suffering large losses in which case they will be risk averse (e.g. buy insurance). On the contrary, people accept risk for moderate to large values of p or to avoid certain or highly probable losses. The later case can be illustrated by a situation of trust under pressure or necessity: if the agent finds himself in an environment where he faces an immediate high danger (e.g. a fire) and has to quickly decide whether or not to use an insecure means (e.g. a damaged rope) to get out of this environment, he will chose to take this risk, thus implicitly trusting the insecure means, since it is a better alternative than death.

These studies show that risk attitudes are not determined by the utility function alone. We will not attempt to formally describe and model risk attitudes. Instead we will simply assume that risk attitudes are individual and context dependent, and based on this attempt to describe some elements of the relationship between trust and risk attitudes. For an overview of alternative approaches to utility theory see Luce (2000) [9] for example.

When analysing the relationship between risk and trust, we will limit ourselves to the case of transactions with two possible outcomes, by associating a gain factor $G_s \in [0, \infty]$ to the outcome of a successful transaction and a loss factor $G_f \in [-1, 0]$ to the outcome of a failed transaction. This can be interpreted as saying that a gain on an investment can be arbitrarily large and that the loss can be at most equal to the investment.

A purely rational and risk-neutral (in the sense that it has no particular propensity to take or avoid risks) agent will decide to enter into a transaction as long as the expected utility is positive. Since risk-neutrality means that $u(IG) = IG$, we use an expression for expected gain without a factor I to determine whether the expected utility will be positive or negative. Given an investment I the return will be IG_s in the case of a

successful transaction, and the loss will be IG_f in case the transaction fails. If we denote by p the probability of success, the *expected gain EG* can then be expressed as:

$$\begin{aligned} EG &= pG_s + (1-p)G_f \\ &= p(G_s - G_f) + G_f \end{aligned} \tag{3}$$

The expected value, which is the same as the expected utility in the case of a risk-neutral attitude, resulting from an investment I can in turn be expressed as:

$$EV = EU = I \cdot EG \tag{4}$$

The parameters G_s, G_f and p determine whether the expected gain is positive or negative. If we assume that a transaction failure causes the total investment to be lost, which can be expressed by setting $G_f = -1$, the expected gain EG is equal to $p(G_s + 1) - 1$, as illustrated in Figure 2.
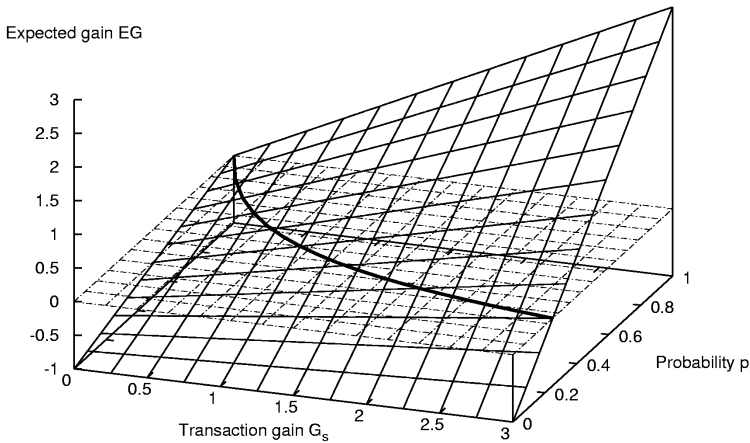
Fig. 2. Expected gain as a function of transaction gain and probability of success.

In Figure 2, the sloping surface (squared and solid line) represents the expected gain for given values of G_s and p, whereas the horizontal surface (squared and dotted line) represents the cut-off surface for zero expected gain. The intersection between the two surfaces is marked with a bold line. Points on the sloping surface above this line represent positive expected gain, whereas points below the line represent negative expected gain.

Figure 2 covers expected gains in the range $[-1, 3]$ (-100% + 300%) but in general the expected gain can be from -1 to any positive value depending on G_s. For example, public lottery can provide gains G_s in the order of several millions although the probability of success is so low that the expected gain is usually negative. An expected gain $EG = -1$ occurs for example when lending money to a con-artist ($p = 0$ and $G_f = -1$).

However, it is not enough to consider whether a transaction has a positive expected gain when making a decision to transact. How much the relying party can afford to loose also plays a role. We examine two examples in order to illustrate this behaviour.

In the first example, an agent deposits amount C in a bank. The transaction consists of keeping the money in the bank for one year in order to earn some interests. The transaction gain G_s is simply the bank's interest rate on savings. The probability of success p is the probability that the money will be secure in the bank. Although money kept in a bank is usually very secure, p can never realistically be equal to 1, so there is a remote possibility that C might be lost. Given that the transaction gain for bank deposits are relatively low, the decision of the relying party to keep his money in the bank can be explained by the perception that there is no safer option easily available.

As another example, let the transaction be to buy a \$1 ticket in a lottery where there are 1,000,000 tickets issued and the price to be won is valued at \$900,000. In this case, according to Equation 3 $G_s = 900,000$, $G_f = -1$ and $p = \frac{1}{1,000,000}$, so that $EG = -\$0.10$. The fact that people still buy lottery tickets can be explained by allocating a value to the thrill of participating in the hope to win the price. By assuming the utility of the thrill to be valued at \$0.11, the expected gain becomes \$0.01, or close to neutral gain. Buying two tickets would not double the thrill and therefore puts the expected gain in negative.

These examples, as well as the case of trust under pressure or necessity previously illustrated, show that people are willing to put different amounts of money at risk depending on the transaction gain and the probability of success. A purely rational agent (in the classic sense) would be willing to invest in any transaction as long as the expected gain is positive. Real people on the other hand will in general not invest all their capital even though the expected gain is positive. More precisely, the higher the probability of success, the higher the fraction of the total capital an agent is willing to put at risk. Let C represent an agent's total capital and $F_C \in [0,1]$ represent the fraction of capital C it is willing to invest in a given transaction. The actual amount I that a person is willing to invest is determined as $I = F_C C$. In the following analysis, we use F_C rather than I because it abstracts the capital value, by normalising the variable that we are studying.

In general F_C varies in the same direction as G_s when p is fixed, and similarly F_C varies like p when G_s fixed. As an example to illustrate this general behaviour let a given agent's risk attitude be determined by the function:

$$F_C(p, G_s) = p^{\frac{\lambda}{G_s}} \tag{5}$$

where $\lambda \in [1, \infty]$ is a factor moderating the influence of the transaction gain G_s on the fraction of total capital that the relying party is willing to put at risk. We will use the term *decision surface* to describe the type of surface illustrated in Figure 3.

λ is interpreted as a factor of the relying party's risk aversion in the given transaction context, and in Fig.3 $\lambda = 10000$. . Independently from the utility function (propensity towards risk), λ represents the contextual component of the risk attitude. A low λ value is representative of a risk-taking behaviour because it increases the volume under the surface delimited by F_C (pushes the decision surface up in Figure 3), whereas a high λ value represents risk aversion because it reduces the volume under the surface (pushes the decision surface down).

Risk attitudes are relative to each individual, so the shape of the surface in Figure 3 only represents an example and will differ for each agent. In this example, we assumed a relatively complex function to represent a non-linear investment behaviour. We do
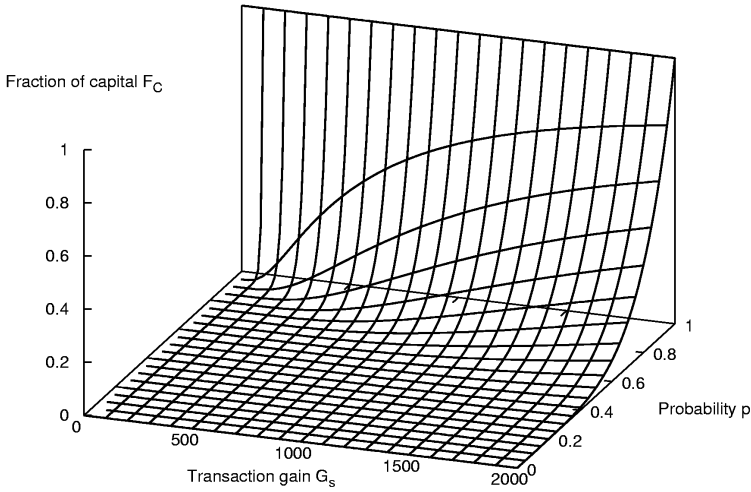
Fig. 3. Example of an agent's risk attitude expressed as a decision surface.

not address here the issue of user modelling, but simply choose a non-trivial example. The surface shape depends as much on the personal preferences of the agent as on its mood in the particular context, but this does not preclude that in unusual situations agent may behave out of the norm, even irrationally. The risk attitude also depends on the total capital C an agent possesses and can change as a function of past experience, notably via the agent's confidence. As already mentioned, we will not try to define general expressions for individual risk attitudes. The expression of Equation 5, with the corresponding surface in Figure 3, only illustrates an example.

A particular transaction will be represented by a point in the 3D space of Figure 3 with coordinates (G_s, p, F_C). Because the surface represents an agent's risk attitude the agent will per definition accept a transaction for which the point is located underneath the decision surface, and will reject a transaction for which the point is located above the decision surface.

3   Balancing Trust and Risk

We now move our point of view on the situation from risk to trust. Whereas in Section 2 the situation was modelled as a transaction, here it revolves around the concepts of dependence and uncertainty. By this we mean that the outcome of the transaction depends on somebody or something and that the relying party is uncertain about the outcome of the transaction.

We assume that transactions can either be successful or failures and that the outcome of a transaction depends on a party x. Furthermore we will let the uncertainty about the outcome of the transaction be represented by the probability p used in Section 2. We can deduce from these hypotheses that p in fact represents the reliability of x for producing a successful outcome, and that p thereby partly represents the trustworthiness of x.

Definition 2 (Reliability Trust). *Reliability trust is defined as the trusting party's probability estimate p of success of the transaction.*

As shown in Section 2, the specific value of p that will make the relying party enter into a transaction also depends on the transaction gain G_s and the invested fraction of the capital F_C

The idea is that, for all combination of values of G_s, p and F_C underneath the decision surface in Figure 3, the relying party trusts x, whereas values above the decision surface lead the relying party to distrust x for this particular transaction. The degree to which the relying party trusts x depends on the distance from the current situation to the decision surface. For example, in the case where G_s is close to zero and F_C is close to one, the relying party will normally not trust x even if p (i.e. the reliability trust) is high.

Since in reality p represents a relative measure of trust and that even agents with high p values can be distrusted, the question is whether it would be useful to determine a better measure of trust, i.e. one that actually measures whether x is trusted for a particular transaction in a given context. Such a measure must necessarily be more complex because of its dependence on gains, investment values and possibly other context-dependent parameters. Although it strictly speaking constitutes an abuse of language to interpret p as a measure of trust, it is commonly being done in the literature. We will therefore not dismiss this interpretation of p as trust, but rather explicitly use the term *reliability trust* to describe it.

Another question which arises when interpreting p as trust is whether it would be better to simply use the concepts of reliability or outcome probability for modelling choice because trust does not add any new information. In fact it has been claimed that the concept of trust is void of semantic meaning in economic theory [17]. We believe that this is an exaggeration and that the notion of trust carries important semantics. The concept of trust is particularly useful in a context of relative uncertainty where a relying party depends on another party to avoid harm and to achieve a successful outcome.

As an attempt to define a measure that adequately represents trusting decisions, we propose to use the normalized difference between x's reliability p and the cut-off probability on an agent's decision surface, what we will call *decision trust*.

Definition 3 (Decision Trust). *Let us assume that: 1) the relying party's risk attitude is defined by a specific decision surface D; 2) a transaction X with party x is characterised by the probability p of x producing a successful outcome, by the transaction gain G_s, and by the fraction of the relying party's capital F_C to be invested in the transaction; 3) p_D is the cut-off probability on the decision surface D for the same values of G_s and F_C. The decision trust T, where $T \in [-1, 1]$, is then defined as:*

$$
\begin{cases}
\text{For } p < p_D : T = \frac{p - p_D}{p_D} \\
\text{For } p = p_D : T = 0 \\
\text{For } p > p_D : T = \frac{p - p_D}{1 - p_D}
\end{cases}
\tag{6}
$$

This decision trust is first defined by its three extreme points: $(0, -1)$, $(p_D, 0)$, and $(1, 1)$. The next constraint is that the decision trust must explicitly depend on a distance

between the current probability p and the cut-off probability p_D. We then choose the most simple functions, given that we have no *a priori* knowledge or experimental data, i.e. a linear function from the distance $\delta = p - p_D$.

A positive decision trust is interpreted as saying that the relying party trusts x for this transaction in this context. A zero decision trust is interpreted as saying that the relying party is undecided as to whether he or she trusts x for this transaction because x's reliability trust is at the cut-off value on the decision surface. Finally, a negative decision trust corresponds to the relying party not trusting x for the transaction in this context.

As an example, Figure 4 illustrates the case of two possible transactions X_1 and X_2. This figure is a section of the decision surface D in Figure 3 for a given value of G_s. The probability difference δ is illustrated for the two transactions X_1 and X_2, as δ_1 and δ_2 respectively. The figure illustrates the case of positive (T_1) and negative (T_2) decision trust, although the actual transaction probability p (i.e. reliability trust) is the same for both situations.
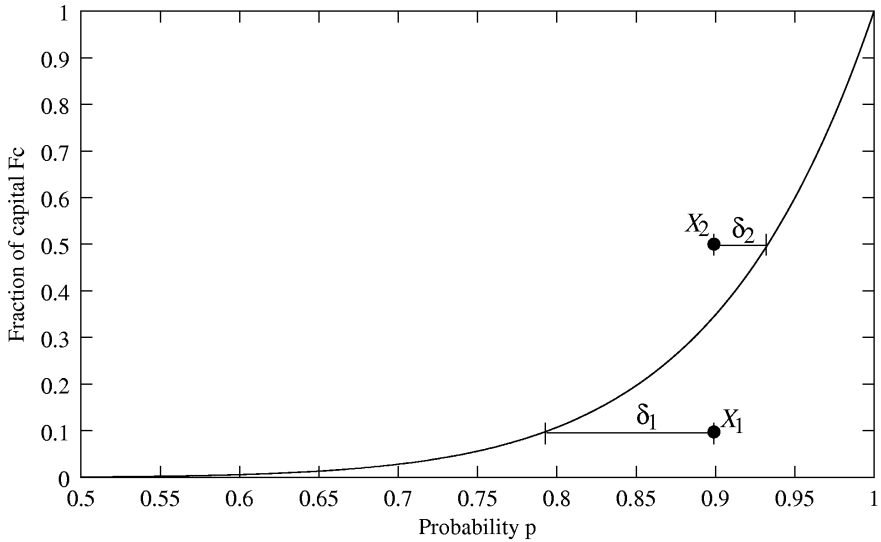
Fig. 4. Illustrating decision trust by difference between reliability and decision cut-off probability.

Finally, one can ask what the relying party should do in the case when the decision trust is negative. Povey (1999) [14] argues that, if the trusting decision is not made, the relying party can treat the risk by: 1) adding countermeasures; 2) deferring risk; 3) or re-computing trust with more or better metrics. Options 1 and 2 aim at increasing the reliability trust p by, respectively, increasing the cost I for the transacting opponent, and hoping that time will soften the relying party's investment. Option 3 confirms our idea

that complex trust measures which are introduced early may sap an interaction. The two interacting parties would then need to re-negotiate the terms of the interaction.

Traditional risk management [15,3] provides us with even more solutions, like risk diversification or risk control, but that falls outside the subject of this study.

4   Conclusion

Risk and trust are two facets of decision-making through which we view the world and choose to act. In an attempt to shape the relationship between risk and trust, this paper tries to refine Manchala's model in order to derive a computational model integrating the two notions. We first compute the transaction's expected gain and illustrate it on several examples. But this transaction factor is not enough to determine the choice of whether to transact or not. We complete this model by introducing the fraction of the capital that an agent is willing to risk.

The intuitive parallel with trust of the first part of our approach is to use the probability of success of the transaction as a measure of trust, what we called *reliability trust*. The decision surface which defines an agent's risk attitude is then taken into account in order to derive a more complete definition of trust, the *decision trust*. This approach provides a more meaningful notion of trust because it combines trust with risk attitudes.

This work is a first step at integrating the two important aspects of decision-making that are risk and trust. We explored their relationship by trying to define a model that could be applied to various examples. Although there is no universal mathematical definition of several aspects of our model (utility function, decision surface) [16], we showed how agent's risk attitudes can be modelled and evaluated in the case of a particular transaction.

As further work, the model needs to be tested with various utility function shapes and decision surfaces, and extended to cope with multiple outcomes. Several other variables can also be integrated into this model. First, we could incorporate more economics and legal information. For example, insurances would consider contractual transactions or the influence of the law, and they could take the form of basic outcomes, minimal investments and specific risk thresholds. Secondly, the temporal aspects should be explored, for example via reinforcement learning or planning techniques to model how an agent adapts to a sequence of transactions. This research activity should tie together trust and risk dynamics. As a continuation of this work, research will be conducted to analyse the situation where trust decision is not made after risk evaluation.

References

1. Bachmann, R. Trust, Power and Control in Trans-Organizational Relations. *Organization Studies*, 2:341–369, 2001.
2. M.H. Birnbaum. Decision and Choice: Paradoxes of Choice. In N.J. Smelser et al., editors, *International Encyclopedia of the Social and Behavioral Sciences*, pages 3286–3291. Elsevier, 2001.
3. D. Borge. *The Book of Risk*. John Wiley & Sons, 2001.
4. Cahill, V. and al. Using Trust for Secure Collaboration in Uncertain Environment. *IEEE Pervasive Computing*, 2(3):52–61, July-September 2003.

5. T. Dimitrakos. A Service-Oriented Trust Management Framework. In R. Falcone, S. Barber, L. Korba, and M. Singh, editors, *Trust, Reputation, and Security: Theories and Practice*, LNAI 2631, pages 53–72. Springer, 2002.
6. R. Falcone and C. Castelfranchi. *Social Trust: A Cognitive Approach*, pages 55–99. Kluwer, 2001.
7. T Grandison. *Trust Management for Internet Applications*. PhD thesis, University of London, July 2003.
8. Grandison, T. and Sloman, M. A Survey of Trust in Internet Applications. *IEEE Communications Surveys*, 3(4):2–16, Fourth Quarter 2000.
9. R.D. Luce. *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. New Jersey: Lawrence Erlbaum Associates, Inc., 2000.
10. D.W. Manchala. Trust Metrics, Models and Protocols for Electronic Commerce Transactions. In *Proc. of the 18th International Conference on Distributed Computing Systems*, pages 312–321. IEEE Computer Society, 1998.
11. A. Mas-Colell, M. Whinston, and J. Green. *Microeconomic Theory*. Oxford University Press, 1995.
12. D. H. McKnight and N. L. Chervany. The Meanings of Trust. Technical Report MISRC Working Paper Series 96-04, University of Minnesota, Management Information Systems Reseach Center, 1996. http://www.misrc.umn.edu/wpaper/wp96-04.htm.
13. Patrick, A. Building Trustworthy Software Agents. *IEEE Internet Computing*, 6(6):46–53, November-December 2002.
14. Povey, D. Developing Electronic Trust Policies Using a Risk Management Model. In *Proc. of the Secure Networking - CQRE (Secure)'99, International Exhibition and Congress*, LNCS 1740, pages 1–16, Düsseldorf, Germany, November 30 - December 2 1999. Springer.
15. E. J. Vaughan. *Risk Management*. John Wiley & Sons, 1997.
16. D. Vose. *Risk Analysis (2nd edition)*. John Wiley & Sons, 2001.
17. O.E. Williamson. *The Economic Institutions of Capitalism*. The Free Press, New York, 1985.