Agents, Information and Trust

John Debenham[1] and Carles Sierra[2]

[1] Faculty of Information Technology, UTS, NSW, Australia
debenham@it.uts.edu.au
http://www-staff.it.uts.edu.au/~debenham/
[2] IIIA, CSIC, UAB, 08193 Bellaterra, Catalonia, Spain

Abstract. Trust measures the relationship between commitment and perceived execution of contracts, and is the foundation for the confidence that an agent has in signing a contract. Negotiation is an information exchange process as well as a proposal exchange process. A rich decision model for intelligent agents involved in multi issue negotiations is described. The model, grounded on information theory, takes into account the aspects of trust and preference to devise mechanisms to manage dialogues. The model supports the design of agents that aim to take 'informed decisions' taking into account that which they have actually observed.

1 Introduction

A negotiating agent, α, uses ideas from information theory to evaluate its negotiation information. If its opponent, β, communicates information, the value of that communication is the decrease in uncertainty in α's model of β. One measure of this decrease in uncertainty is *Shannon information* [10], or negative entropy. If α communicates information it evaluates that information as its expectation of the resulting decrease in β's uncertainty about α. Any such decrease in uncertainty is seen against the continually increasing uncertainty in information because information integrity necessarily decays in time. Information theory may also be used to measure features of inter-agent relationships that extend beyond single negotiations — for example, the sharing of information in a trading pact or relationship, and the strength of trading networks that form as a result of such information sharing. α uses entropy-based inference — both maximum entropy inference and minimum relative entropy inference — to derive probability distributions for that which it does not know in line with the following principle.

Information Principle. α's information base contains only observed facts — in the absence of observed facts, α may speculate about what those facts might be. For example in competitive negotiation, β's utility, deadlines, and other private information will never be observable — unless β is foolish. Further, β's motivations (such as being a utility optimizer) will also never be observable. So in competitive negotiation β's private information is "off α's radar" — α does not contemplate it or speculate about it.

S. Zhang and R. Jarvis (Eds.): AI 2005, LNAI 3809, pp. 643–652, 2005.

We assume that the interactions between agents are made within the framework of an infrastructure that fixes ontology and meaning, for instance an Electronic Institution [1]. Thus, no differences in the observation of illocutions have to be assumed. Moreover, we assume that the role of a player is public information that can be observed — this is the case for instance if we are within an electronic institution framework. However, the perception of the execution of a contract is subjective in the sense that two deviations of behaviour can be perceived differently by two different agents. We will therefore assume that each agent is equipped with a perception function (noted in this paper as Observe(·)) that determines which contract execution has actually occurred. Trust measures the relationship between commitment and execution of contracts. More precisely, between signed contracts and *perceived* execution of contracts. In this way, a natural way to base our modelling of trust is on a conditional probability, P^t, between contracts given a context e as:

$$P^t(\text{Observe}(\alpha, b')|\text{Accept}(\beta, \alpha, (a, b)), e)$$

where every contract execution represents a point in that distribution.[1] A concrete relation between a signed contract and the perception of an executed one.

2 Trust and Negotiation

Trust is a multi-faceted concept that has received increasing attention recently [11,12,13,6]. In the context of negotiation, trust represents a general assessment on how 'serious' an agent is about the negotiation process, i.e. that his proposals 'make sense' and he is not 'flying a kite', and that he is committed to what he signs. A lack of trust may provoke agents to breakdown negotiations, or to demand additional guarantees to cover the risk of potential defections. Therefore, in any model of trust the central issue is how to model expectations about the actual outcome at contract execution time. Contracts, when executed, may, and frequently do, yield a different result to what was initially signed. Goods may be delivered late, quality may be (perceived) different from the contract specification, extra costs may be claimed, etc. So the outcome is uncertain to some extent, and trust, precisely, is a measure of how uncertain the outcome of a contract is. Naturally, the higher the trust in a partner the more sure we are of his or her reliability. Trust is therefore a *measure of expected deviations of behaviour* along a given dimension, and in many cases for a given value (region) in that dimension (e.g. I might trust you on low-priced contracts but not on high-priced ones). In this sense, the higher the trust the lower the expectation that a (significant) deviation from what is signed occurs.

Rhetorics in a negotiation context represents the use of language constructs to persuade the oponent to accept our proposals. Agents use rhetorics because they want to change their opponents' preferences or their opponents' view of

[1] To simplify notation in the rest of the paper we will denote $P^t(\text{Observe}(\alpha, b')|\text{Accept}(\beta, \alpha, (a, b), e)$ as $P^t(b'|b, e)$.

them. Rhetorical constructs are intimately linked to *social structure* and *time*. The *amount* and the *persistence* of preference change induced by a rhetorical construct depends strongly on the distance in some social scale between the speaker and the hearer. The larger the distance the bigger the impact. The closer, the longer the effect. A Nobel Prize winner is able to change *a lot* our views on a certain subject although getting convinced by a peer on some matter usually has *longer* impact [9].

Another dimension that is very important in the analysis of dialogues is ontology. The contents of illocutions determine whether our assumptions about the opponent's model of the problem are correct. New values can be added to a dimension by a simple question: "Have you got yellow plastic crocodiles?". A simplifying solution is to start any negotiation process by fixing a common ontology. Alternatively, we can use ontology clarifying dialogues and then be ready to modify our models during the dialogue.

3 A Negotiation Language

Agent α is negotiating with an opponent β. They aim to strike a deal $\delta = (a, b)$ where a is α's commitment and b is β's. (Where a or b might be empty.) We denote by A the set of all possible commitments by α, and by B the set of all possible commitments by β. The agents have two languages, $\mathcal{C}$ for communication (illocutionary based) and $\mathcal{L}$ for internal representation (as a restricted first-order language).[2]

In this paper we assume that the illocution particle set is:

$$\iota = \{\text{Offer}, \text{Accept}, \text{Reject}, \text{Withdraw}, \text{Inform}\}$$

with the following syntax and informal meaning:

- Offer(α, β, δ) Agent α offers agent β a deal $\delta = (a, b)$ with action commitments a for α and b for β.
- Accept(α, β, δ) Agent α accepts agent β's previously offered deal δ.
- Reject$(\alpha, \beta, \delta, [\mathit{info}])$ Agent α rejects agent β's previously offered deal δ. Optionally, information explaining the reason for the rejection can be given.
- Withdraw$(\alpha, \beta, [\mathit{info}])$ Agent α breaks down negotiation with β. Extra *info* justifying the withdrawal may be given here.
- Inform$(\alpha, \beta, [\mathit{info}])$ Agent α informs β about *info*.

The accompanying information, *info*, can be of two basic types: (i) referring to the process (plan) used by an agent to solve a problem, or (ii) data (beliefs) of the agent including preferences. When negotiating, agents will therefore try

[2] It is commonly accepted since the works by Austin and Searle that illocutionary acts are actions that succeed or fail. We will abuse notation in this paper and will consider that they are predicates in a first order logic meaning 'the action has been performed'. For those more pure-minded an alternative is to consider dynamic logic.

to influence the opponent by trying to change their processes (plans) or by providing new data.

Following the extensive literature on preferences, preferences are divided into two classes:

Quantitative. These preferences are usually called *soft constraints* (hard constraints are particular cases of soft constraints). A soft constraint associates each instantiation of its variables with a value from a partially ordered set. One natural interpretation of this value is the probability of choice. In general, preferences can be expressed as values within a semi-ring $\langle A, +, \times, 0, 1\rangle$ such that A is a set and $0, 1 \in A$; $+$ is commutative, associative with 0 as its unit element; $\times$ is associative, distributes over $+$, 1 is its unit element, and 0 is its annihilating element. Given a semi-ring $\langle A, +, \times, 0, 1\rangle$, an ordered set of variables $V = \{v_1, ..., v_n\}$ and their corresponding domains $D = \{D_1, ..., D_n\}$ a soft constraint is a pair $\langle f, con\rangle$ where $con \subseteq V$ and $f \colon D^{|con|} \rightarrow A$ with the following intuitive meaning: $f(d_1, \ldots, d_n) = \kappa$ means that the binding $x_1 = d_1, \ldots, x_n = d_n$ satisfies the constraint to a level of κ.[3]

Qualitative. In many domains it is difficult to formulate precise numerical preferences, and it is more convenient to express preference relations between variable assignments: "I prefer red cars to yellow cars". The usual way to represent this relationship formally is $v = a > v = a'$, or simpler $a > a'$, meaning that we prefer the assignment of variable v to a than to a'. Also, in case of an absolute preference for a particular value in a domain, that is, when our preference is $\forall x \neq a.v = a > v = x$ we can simply write $v = a$ or just a. Also, in many cases preference relations depend on the values assigned to other variables (configuring what is called a Conditional Preference Net (CP-net) [2]). "If they serve meat, I prefer red wine to white wine". A conditional preference can be represented as $v_1 = c : v_2 = d > v_2 = d'$ (or again $c : d > d'$) meaning that we prefer d to d' in the context where c is the case. In general, any DNF over value assignments could be used as the condition. And also, other comparatives than '=' could be used.

Finally, it seems natural that the constraints have an associated certainty degree representing their degree of truth. We thus propose the following content language expressed in BNF: ($info \in \mathcal{L}$):

$$\begin{array}{ll}
\mathit{info} & ::= \mathit{unit}[\ \textbf{and}\ \mathit{info}] \\
\mathit{unit} & ::= K | B | \mathit{soft} | \mathit{qual} | \mathit{cond} \\
K & ::= \mathbf{K(}\,\mathit{WFF}\mathbf{)} \\
B & ::= \mathbf{B(}\,\mathit{WFF}\mathbf{)} \\
\mathit{soft} & ::= \mathbf{soft(}f, \{V^+\}\mathbf{)}
\end{array}$$

[3] As we use maximum entropy inference we have to make the simplifying assumption that domains of quantitative constraints must be finite. This means that continuous domains must be represented as a finite set of intervals, further the way in which those intervals are chosen affects the outcome. This is sometime cited as a weakness of the maximum entropy approach. In [3] it is argued to the contrary, that the choice of intervals represents our prior expectations in fine-grained detail.

qual ::= *V*=*D*[>*V*=*D*]
cond ::= **If** *DNF* **Then** *qual*
WFF ::= *any wff over subsets of variables* $\{V\}$
DNF ::= *conjunction*[**or** *DNF*]
conjunction::= *qual*[**and** *conjunction*]
V ::= $\mathbf{v_1} | \cdots | \mathbf{v_n}$
D ::= **a**|**a'**|**b**| $\cdots$
f ::= *any function from the domains of subsets of V to a set A. For instance a fuzzy set membership function if A = [0,1]*

4 Information-Based Negotiation

We ground our negotiation model on information-based concepts. *Entropy*, H, is a measure of uncertainty [10] in a probability distribution for a discrete random variable X: $H(X) \triangleq -\sum_i p(x_i)\log p(x_i)$ where $p(x_i) = P(X = x_i)$. Maximum entropy inference is used to derive sentence probabilities for that which is not known by constructing the “maximally noncommittal” [8] probability distribution.

Let $\mathcal{G}$ be the set of all positive ground literals that can be constructed using our language $\mathcal{L}$. A *possible world*, v, is a valuation function: $\mathcal{G} \rightarrow \{\top, \bot\}$. $\mathcal{V}|\mathcal{K} = \{v_i\}$ is the set of all possible worlds that are consistent with an agent's knowledge base $\mathcal{K}$ that contains statements which the agent believes are true. A *random world* for $\mathcal{K}$, $W|\mathcal{K} = \{p_i\}$ is a probability distribution over $\mathcal{V}|\mathcal{K}^a = \{v_i\}$, where p_i expresses an agent's degree of belief that each of the possible worlds, v_i, is the actual world. The *derived sentence probability* of any $\sigma \in \mathcal{L}$, *with respect to* a random world $W|\mathcal{K}$ is:

$$(\forall \sigma \in \mathcal{L}) P_{\{W|\mathcal{K}\}}(\sigma) \triangleq \sum_n \{\, p_n \;:\; \sigma\, is \top in\, v_n \,\} \tag{1}$$

The agent's *belief set* $\mathcal{B} = \{\varphi_j\}_{j=1}^M$ contains statements to which the agent attaches a *given sentence probability* $B(\cdot)$. A random world $W|\mathcal{K}$ is *consistent* with $\mathcal{B}$ if: $(\forall \varphi \in \mathcal{B})(B(\varphi) = P_{\{W|\mathcal{K}\}}(\varphi))$. Let $\{p_i\} = \{\overline{W}|\mathcal{K}, \mathcal{B}\}$ be the “maximum entropy probability distribution over $\mathcal{V}|\mathcal{K}$ that is consistent with $\mathcal{B}$”. Given an agent with $\mathcal{K}$ and $\mathcal{B}$, *maximum entropy inference* states that the *derived sentence probability* for any sentence, $\sigma \in \mathcal{L}$, is:

$$(\forall \sigma \in \mathcal{L}) P_{\{\overline{W}|\mathcal{K},\mathcal{B}\}}(\sigma) \triangleq \sum_n \{\, p_n \;:\; \sigma\, is \top in\, v_n \,\} \tag{2}$$

From Eqn. 2, each belief imposes a linear constraint on the $\{p_i\}$. The maximum entropy distribution: $\arg\max_{\underline{p}} H(\underline{p})$, $\underline{p} = (p_1, \ldots, p_N)$, subject to $M + 1$ linear constraints:

$$g_j(\underline{p}) = \sum_{i=1}^{N} c_{ji} p_i - B(\varphi_j) = 0, \quad j = 1, \ldots, M. \quad g_0(\underline{p}) = \sum_{i=1}^{N} p_i - 1 = 0$$

$c_{ji} = 1$ if φ_j is $\top$ in v_i and 0 otherwise, and $p_i \geq 0, i = 1, \ldots, N$, is found by introducing Lagrange multipliers, and then obtaining a numerical solution using the multivariate Newton-Raphson method. In the subsequent subsections we'll see how an agent updates the sentence probabilities depending on the type of information used in the update.

An important aspect that we want to model is the fact that beliefs 'evaporate' as time goes by. If we don't keep an ongoing relationship, we somehow forget how *good* the opponent was. If I stop buying from my butcher, I'm not sure anymore that he will sell me the 'best' meat. This decay is what justifies a continuous relationship between individuals. In our model, the conditional probabilities should tend to ignorance. If we have the set of observable contracts as $B = \{b_1, b_2, \ldots, b_n\}$ then complete ignorance of the opponent's expected behaviour means that given the opponent commits to b the conditional probability for each observable contract becomes $\frac{1}{n}$ — i.e. the unconstrained maximum entropy distribution. This natural decay of belief is offset by new observations.

We define the evolution of the probability distribution that supports the previous definition of decay using an equation inspired by pheromone like models [5]:

$$P^{t+1}(b'|b) = \kappa \cdot \left(\frac{1-\rho}{n} + \rho \cdot \left(P^t(b'|b) + \Delta^t P(b'|b)\right) \right) \tag{3}$$

where κ is a normalisation constant to ensure that the resulting values for $P^{t+1}(b'|b)$ are a probability distribution. This equation models the passage of time for a conveniently large $\rho \in [0, 1]$ and where the term $\Delta^t P(b'|b)$ represents the increment in an instant of time according to the last experienced event as the following possibilities show.

Similarity based. The question is how to use the observation of a contract execution c' given a signed contract c in the update of the overall probability distribution over the set of all possible contracts. Here we use the idea that given a particular deviation in a region of the space, *similar* deviations should be expected in other regions. The intuition behind the update is that if my butcher has not given me the quality that I expected when I bought lamb chops, then I might expect similar deviations with respect to chicken. This idea is built upon a function $f(x, y)$ that takes into account the difference between acceptance probabilities and similarity between the perception of the execution x of a contract y, that is a contract for which there was an $\text{Accept}(\beta, \alpha, y)$. Thus, after the observation of c' the increment of probability distribution at time $t + 1$ is:

$$\Delta^t P(b'|b) = (1 - |f(c', c) - f(b', b)|) \tag{4}$$

where $f(x, y)$ is

$$f(x, y) = \begin{cases} 1 & \text{if } P^t(\text{Accept}(x)) > P^t(\text{Accept}(y)) \\ \text{Sim}(x, y) & \text{otherwise.} \end{cases}$$

and where Sim is an appropriate similarity function (reflexive and symmetric) that determines the indistinguishability between the perceived and the committed contract.

Entropy based. Suppose that outcome space is $B = (b_1, \ldots, b_m)$, then for a given b_k, $(P^t_\beta(b_1|b_k), \ldots, P^t_\beta(b_m|b_k))$ is the prior distribution of α's estimate of what β will actually execute if he contracted to deliver b_k. Suppose that α's evaluation space is $E = (e_1, \ldots, e_n)$ with evaluation function $\underline{v}$, then $\underline{v}(b_k) = (v_1(b_k), \ldots, v_n(b_k))$ is α's evaluation over E of what β contracted to do. Then α's expected evaluation of what β will deliver if β contracts to deliver b_k is:

$$\underline{v}_\beta(b_k) = \left(\sum_{j=1}^{m} P^t_\beta(b_j|b_k) \cdot v_1(b_j), \ldots, \sum_{j=1}^{m} P^t_\beta(b_j|b_k) \cdot v_n(b_j) \right).$$

Now suppose that α observes the event $(c'|c)$, α may wish to revise the prior estimate $\underline{v}_\beta(b)$ in the light of this observation to:

$$(\underline{v}_\beta(b) \mid (c'|c)) = \underline{g}(\underline{v}_\beta(b), \underline{v}(c), \underline{v}(c')),$$

for some function $\underline{g}$ — the idea being, for example, that if the chicken, c', was tough then our expectation that the beef, b', will be tough should increase. The entropy based approach achieves this by estimating $\Delta^t P(b'|b)$ by applying the principle of minimum relative entropy — see Sec. 4. Let:

$$\left(P^t_{\beta,C}(b_j|b)\right)_{j=1}^{n} = \arg\min_{\underline{p}} \sum_{i=1}^{n} p_i \log \frac{p_i}{P^t_\beta(b_i|b)} \tag{5}$$

satisfying the n constraints C, and $\underline{p} = (p_j)_{j=1}^n$. Then:

$$\Delta^t P(b'|b) = P^t_{\beta,C}(b'|b) - P^t_\beta(b'|b) \tag{6}$$

The n constraints C are: $\sum_{j=1}^{m} P^t_\beta(b_j|b_k) \cdot v_i(b_j) = g_i(\underline{v}_\beta(b), \underline{v}(c), \underline{v}(c'))$, for $i = 1, \ldots, n$. This is a set of n linear equations in m unknowns, and so the calculation of the minimum relative entropy distribution may be impossible if $n > m$. In this case, we take only the m equations for which the change from the prior to the posterior value is greatest. That is, we attempt to select the most significant factors.

5 A Trust Model

"Trust" may have different significance in different contexts. Agents build and destroy trust by the way in which they execute their contractual commitments. If agent α who is committed to execute a actually executes a' then we distinguish two ways in which a and a' may differ. First, a' may be a variation of commitment a within the ontological context of the negotiation. Second, the contract variation may involve something outside the ontological context. A contract execution could involve variations of both of these types. In the following we are primarily interested in variations of the first type. Variations of the second type can be managed by reference to market data that together with entropy-based inference enables α to value such variations.

Trust as expected behaviour. Consider a distribution of expected contract executions that represent α's "ideal" for β, in the sense that it is the best that α could reasonably expect β to do. This distribution will be a function of β, α's trading history with β, anything else that α believes about β, and general environmental information including time — denote all of this by e, then we have $P^t_I(b'|b,e)$. For example, if it is unacceptable for the execution b' to be less preferred than the agreement b then $P^t_I(b'|b,e)$ will only be non-zero for those b' that α prefers to b. The distribution $P^t_I(\cdot)$ represents what α expects, or hopes, β will do. So if β commits to deliver 12 bottles of water then the probability of b' being a case of champagne could be low. Trust is the relative entropy between this ideal distribution, $P^t_I(b'|b,e)$, and the distribution of the observation of expected contract execution, $P^t_\beta(b'|b)$. That is:

$$T(\alpha,\beta,b) = 1 - \sum_{b' \in B(b)} P^t_I(b'|b,e) \log \frac{P^t_I(b'|b,e)}{P^t_\beta(b'|b)}$$

This defines trust for one, single commitment b — for example, my trust in my butcher if I order precisely seventeen lamb loin chops. It makes sense to aggregate these values over a class of commitments, say over those b' that satisfy some first-order formula $\Phi(\cdot)$. In this way we measure the trust that I have in my butcher for lamb cuts generally:

$$T(\alpha,\beta,\Phi) = 1 - \frac{\sum_{\{b \in B|\Phi(b)\}} P^t_\beta(b) \left[\sum_{b' \in B(b)} P^t_I(b'|b,e) \log \frac{P^t_I(b'|b,e)}{P^t_\beta(b'|b)}\right]}{\sum_{\{b \in B|\Phi(b)\}} P^t(b)}$$

where $P^t(b)$ is the probability of α signing a contract with β that involves the commitment b. Similarly, for an overall estimate of α's trust in β:

$$T(\alpha,\beta) = 1 - \sum_{b \in B} P^t(b) \left[\sum_{b' \in B(b)} P^t_I(b'|b,e) \log \frac{P^t_I(b'|b,e)}{P^t_\beta(b'|b)}\right]$$

Trust as expected acceptability. The notion of trust in the previous section Sec. 5 was expressed in terms of our expected behaviour in an opponent that was defined for each contract specification b. That notion requires that an ideal distribution, $P^t_I(b'|b,e)$, has to be specified for each b. The specification of ideal distributions may be avoided by considering "expected acceptability" instead of "expected behaviour". The idea is that we trust β if the acceptability of his contract executions are at or marginally above the acceptability of the contract specification, $P^t(\text{IAcc}(\alpha,\beta,\nu,(a,b)))$. Defining a function:

$$f(x) = \begin{cases} 0 & \text{if } \; x < P^t(\text{IAcc}(\alpha,\beta,\nu,(a,b))) \\ 1 & \text{if } \; P^t(\text{IAcc}(\alpha,\beta,\nu,(a,b))) < x < P^t(\text{IAcc}(\alpha,\beta,\nu,(a,b))) + \epsilon \\ 0 & \text{otherwise} \end{cases}$$

(or perhaps a similar function with smoother shape) for some small ϵ, then define:

$$T(\alpha,\beta,b) = \sum_{b' \in B} f(P^t(\mathrm{IAcc}(\alpha,\beta,\nu,(a,b')))) \cdot P^t_\beta(b'|b)$$

$$T(\alpha,\beta,\Phi) = \frac{\sum_{\{b \in B | \Phi(b)\}} P^t(b) \cdot \sum_{b' \in B} f(P^t(\mathrm{IAcc}(\alpha,\beta,\nu,(a,b')))) \cdot P^t_\beta(b'|b)}{\sum_{\{b \in B | \Phi(b)\}} P^t(b)}$$

$$T(\alpha,\beta) = \sum_{b \in B} P^t(b) \cdot \sum_{b' \in B} f(P^t(\mathrm{IAcc}(\alpha,\beta,\nu,(a,b')))) \cdot P^t_\beta(b'|b)$$

Trust as certainty in contract execution. Trust is consistency in expected acceptable contract executions, or "the lack of expected uncertainty in those possible executions that are better than the contract specification". The idea here is that α will trust β more if variations, b', from expectation, b, are not random. The Trust that an agent α has on agent β with respect to the fulfilment of a contract (a,b) is:

$$T(\alpha,\beta,b) = 1 + \frac{1}{B^*} \cdot \sum_{b' \in B} P^t_+(b'|b) \log P^t_+(b'|b)$$

where $P^t_+(b'|b)$ is the normalisation of $P^t_\beta(b'|b)$ for those values of b' for which $P^t(\mathrm{IAcc}(\alpha,\beta,\nu,(a,b'))) > P^t(\mathrm{IAcc}(\alpha,\beta,\nu,(a,b)))$ and zero otherwise, $B(b)^+$ is the set of contract executions that α prefers to b,

$$B^* = \begin{cases} 1 & \text{if } |B(b)^+| = 1 \\ \log |B(b)^+| & \text{otherwise} \end{cases}$$

and β has agreed to execute b, and α systematically observes b'. Given some b' that α does not prefer to b, the trust value will be 0. Trust will tend to 0 when the dispersion of observations is maximal.

As above we aggregate this measure for those deals of a particular type, that is, those that satisfy $\Phi(\cdot)$:

$$T(\alpha,\beta,\Phi) = 1 + \frac{\sum_{\{b \in B | \Phi(b)\}} \sum_{b' \in B} \left[P^t_+(b',b) \log P^t_+(b'|b) \right]}{B^* \cdot \sum_{\{b \in B | \Phi(b)\}} P^t(b)}$$

where $P^t_\beta(b',b)$ is the joint probability distribution. And, as a general measure of α's trust on β we naturally use the normalised negative conditional entropy of executed contracts given signed contracts:

$$T(\alpha,\beta) = 1 + \frac{\sum_{b \in B} \sum_{b' \in B} \left[P^t_+(b',b) \log P^t_+(b'|b) \right]}{B^*}$$

6 Conclusion

Game theory tells α that she should accept a proposal if $s_\delta > m_\delta$ where s_δ is the surplus, $s_\delta = u(\omega) - u(\pi)$ and m_δ is the margin. This is fine if everything is certain. If it is not then game theory tells α to work with a random variable, S_δ, instead. Incidentally this means that α has to be certain about her uncertainty, but that is not the issue. This means that α can consider $P(S_\delta > m_\delta)$, and the standard deviation, $\sigma(S_\delta)$, is a measure of uncertainty in the process. Then α asks "how risk averse am I", and then is able to calculate $P((accept(\delta))$.

With uncertain information and decaying integrity, the "utility calculation" in the previous paragraph is a futile exercise. Instead we argue that it makes more sense to ask simply: "on the basis of what we actually know, what is the best thing to do?". We claim that α will be more concerned about the integrity of the information with which the decision is being made, than with an uncertain estimation of her utility distribution.

References

1. J. L. Arcos, M. Esteva, P. Noriega, J. A. Rodríguez, and C. Sierra. Environment engineering for multiagent systems. *Journal on Engineering Applications of Artificial Intelligence*, 18, 2005.
2. C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135 – 191, 2004.
3. J. Debenham. Auctions and bidding with information. In P. Faratin and J. Rodriguez-Aguilar, editors, *Proceedings Agent-Mediated Electronic Commerce VI: AMEC*, pages 15 – 28, July 2004.
4. J. Debenham. Bargaining with information. In N. Jennings, C. Sierra, L. Sonenberg, and M. Tambe, editors, *Proceedings Third International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2004*, pages 664 – 671. ACM, July 2004.
5. M. Dorigo and T. Stützle. *Ant Colony Optimization.* MIT Press, Cambridge, MA, 2004.
6. R. Falcone and C. Castelfranchi. The socio-cognitive dynamics of trust: Does trust create trust? In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies*, pages 55 – 72, 2001.
7. P. Faratin, C. Sierra, and N. Jennings. Using similarity criteria to make issue tradeoffs in automated negotiation. *Journal of Artificial Intelligence*, 142(2):205–237, 2003.
8. E. Jaynes. *Probability Theory — The Logic of Science.* Cambridge University Press, 2003.
9. M. Karlins and H. Abelson. *Persuasion.* Crosby Lockwood and Son, 1970.
10. D. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.
11. S. Ramchurn, C. Sierra, L. Godo, and N. Jennings. Devising a trust model for multiagent interactions using confidence and reputation. *International Journal of Applied Artificial Intelligence*, 18(9–10):91–204, 2005.
12. J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the First International Conference on Autonomous Agents and Multi-Agent systems*, pages 475 – 482, 2002.
13. P. Yolum and M. Singh. Achieving trust via service graphs. In *Proceedings of the Autonomous Agents and Multi-Agent Systems Workshop on Deception, Fraud and Trust in Agent Societies.* Springer-Verlag, 2003.