Information-based Deliberation

Carles Sierra
Institut d'Investigacio en Intel.ligencia Artificial
Spanish Scientific Research Council, UAB
08193 Bellaterra, Catalonia, Spain
sierra@iiia.csic.es

John Debenham
Faculty of Information Technology
University of Technology, Sydney
NSW, Australia
debenham@it.uts.edu.au

ABSTRACT
Information-based agency is founded on two observations: everything in an agent's world model is uncertain, and everything that an agent communicates gives away valuable information. The agent's deliberative mechanism manages interaction using plans and strategies in the context of the relationships the agent has with other agents, and is the means by which those relationships develop.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms
Theory

Keywords
Multiagent systems, Negotiation

1. INTRODUCTION
This paper is in the area labelled: *information-based agency* [14]. An information-based agent has an identity, values, needs, plans and strategies all of which are expressed using a fixed ontology in probabilistic logic for internal representation and in an illocutionary language [13] for communication. All of the forgoing is represented in the agent's deliberative machinery. We assume that such an agent resides in a electronic institution [1] and is aware of the prevailing norms and interaction protocols. In line with our "Information Principle" [12], an information-based agent makes no *a priori* assumptions about the states of the world or the other agents in it — these are represented in a world model, $\mathcal{M}^t$, that is inferred solely from the messages that it receives.

The world model, $\mathcal{M}^t$, is a set of probability distributions for a set of random variables each of which represents the agent's expectations about some point of interest about the world or the other agents in it. We build a history of interaction by noting each commitment made (commitments to act, commitments to the truth of information or to the validity of an argument), and by relating each of them to subsequent observations of what occurs. Tools from information theory are then used to summarise these historic (commitment, observation) pairs — in this way we have defined models of *trust*, *honour*, *reliability* and *reputation* [13]. Further we have defined the *intimacy* and *balance* of both dialogues and relationships [15] in terms of our 'LOGIC' illocutionary framework. All of these notions make no presumption that our agents will align themselves with any particular strategy.

The information-based approach is concerned with modelling the speaker by observing the *interaction process*. In contrast the utilitarian approach is primarily concerned with *outcomes*, in particular with attempting to achieve the most preferred outcome. An information-based agent may operate as a utility optimiser if its preferences are known outright — e.g. if its preferences are expressed in terms of the ontology independent of any world or agent states. For example, it may prefer to pay less than pay more. If its preferences involve world or agent states then they will be modelled as probability distributions in $\mathcal{M}^t$, where they can support a utilitarian strategy provided that the agent's belief in them is sufficiently strong.

We have described argumentation strategies. For example, the *equitable information gain* strategy attempts to reply to an utterance with a response that will give the opponent expected information gain that is similar to that which the agent observed when the utterance was received. This strategy may be used in a rich argumentation setting to assist an agent to choose from a set of possible responses. In the LOGIC negotiation framework [15] we show how *inequitable* information gain in the responses may be used to gradually develop a relationship over repeated interaction rounds. Estimates of information gain across five classes of illocutionary acts form the basis for an on-going relationship-building strategy that attempts to move the relationship towards a 'relationship target'. It is guided by estimates of *intimacy* (the current degree of 'closeness' in the relationship) and of *balance* (the observed degree of 'fairness' in the exchanges).

In related papers we have focused on argumentation strategies, trust and honour, and have simply assumed that the agent has a kernel deliberative system. In this paper we describe the deliberative system for an information-based agent, and show how it may accommodate utilitarian thinking if required.

Information-based agency is founded on two premises. First, everything in its world model is uncertain. Second, everything that an agent communicates gives away valuable

Cite as: Information-based Deliberation, Carles Sierra & John Debenham, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons(eds.),May,12-16.,2008,Estoril,Portugal,pp. 689-696.

information. Information, including arguments, may have no particular utilitarian value, and so may not readily be accommodated by an agent's utilitarian machinery. We argue that information should not only be valued in terms of the agent's preferences. We discuss uncertainty in Section 2. Section 3 discusses a utilitarian approach and an information approach to valuing information. Section 4 describes the kernel of the deliberative mechanism, that is employed in a discussion of strategies in Section 5, and Section 6 concludes.

2. UNCERTAINTY

This discussion is from the point of view of an information-based agent α in a multiagent system where α interacts with negotiating agents, β_i, information providing agents, θ_j, and an *institutional agent*, ξ, that represents the institution where we assume that all interactions happen [1]: $\{\alpha, \beta_1, \dots, \beta_o, \xi, \theta_1, \dots, \theta_t\}$.

A pair of agents interact by passing messages to each other. We assume that they share a common ontology and that their interactions are organised into dialogues, where a *dialogue* is a finite sequence of inter-related utterances. A *commitment* is a consequence of an utterance by an agent that contains a promise that the world will be in some state in the future. A *contract* is a pair of commitments exchanged between a pair[1] of agents. The set of all dialogues between two agents up to the present is their *relationship*.

We represent commitments using conditional probabilities, $\mathbb{P}(\varphi'|\varphi)$. If φ is a commitment that is expected to be fulfilled to some degree, and φ' the corresponding subsequent enactment then $\mathbb{P}(\varphi'|\varphi)$ is the probability that φ' will be observed given that φ was expected. For example, if φ is a commitment signed by β then the conditional probability, $\mathbb{P}(\varphi'|\varphi)$, is an estimate of α's expectation of β's execution of that commitment, and the uncertainty in β's execution of his commitments is the entropy $\mathbb{H}(\varphi'|\varphi)$.

In a multiagent system it is natural to measure the uncertainty of a random variable in terms of the cost, in some sense, of communicating the true value of it from one agent to another. One such sense is the lower bound on the number of binary questions that are always guaranteed to discover the true value of a random variable, X; this is given by the *entropy*, $\mathbb{H}(X) = \sum_i -p_i \log p_i$, where the p_i are values of the probability mass function for X, [10].

Agent α's world model, $\mathcal{M}^t$, at time t is a set of random variables, $\mathcal{M}^t = \{X_i, \dots, X_n\}$ each representing an aspect of the world that α is interested in. In the absence of new information the integrity of $\mathcal{M}^t$ should decrease in time. α may have background knowledge concerning the expected probability mass function for a variable X as $t \to \infty$. Such background knowledge is represented as a *decay limit distribution*. Given a probability mass function, $\mathbb{P}(X_i)$, for variable X_i, and a decay limit distribution $\mathbb{D}(X_i)$: $\mathbb{P}^{t+1}(X_i) = d_i(\mathbb{D}(X_i), \mathbb{P}^t(X_i))$, where d_i is the *decay function* for the X_i satisfying the property that $\lim_{t\to\infty} \mathbb{P}^t(X_i) = \mathbb{D}(X_i)$. Either the decay function or the decay limit distribution could also be a function of time: d_i^t and $\mathbb{D}^t(X_i)$ [14]. For simplicity we use a linear decay function; $\mathbb{P}(X_i)$ decays by:

$$\mathbb{P}^{t+1}(X_i) = \lambda \times \mathbb{D}(X_i) + (1 - \lambda) \times \mathbb{P}^t(X_i) \qquad (1)$$

for some constant λ: $0 < \lambda < 1$, where λ is the *decay rate*.

Suppose that α receives some new information in an utterance u from agent β at time t where u states that something is so with some probability z, and suppose that α attaches an epistemic belief $\mathbb{R}^t(\alpha, \beta, u)$ to u — this probability reflects α's level of personal *caution*. [14] gives a method for estimating $\mathbb{R}^t(\alpha, \beta, u)$ on the basis of observation. [14] also describes how each of α's active plans, s, contains constructors for a set of distributions $\{X_i\} \in \mathcal{M}^t$ together with associated *update functions*, $J_s(\cdot)$, such that $J_s^{X_i}(u)$ is a set of linear constraints on the posterior distribution for X_i. Let $\mathbb{P}^t(X_{i,(u)})$ be the distribution with minimum relative entropy with respect to the prior $\mathbb{P}^t(X_i)$ subject to the constraints $J_s^{X_i}(u)$. Then let $\mathbb{P}^t(\overline{X}_{i,(u)}) =$

$$\mathbb{R}^t(\alpha, \beta, u) \times \mathbb{P}^t(X_{i,(u)}) + (1 - \mathbb{R}^t(\alpha, \beta, u)) \times \mathbb{P}^t(X_{i,(u)})$$

The posterior $\mathbb{P}^{t+1}(X_i)$ is $\mathbb{P}^t(\overline{X}_{i,(u)})$ as long as $\mathbb{P}^t(\overline{X}_{i,(u)})$ is "more interesting[2]" than $\mathbb{P}^t(X_i)$, otherwise it is the prior. This process takes account of both the belief $\mathbb{R}^t(\alpha, \beta, u)$, and the probability z that will be represented in $J_s^{X_i}(u)$. The update functions, J, are the bridge between the illocutionary language and the agent's world model.

In his 1957 paper [5], E.T. Jaynes wrote: "Information theory ... leads to a type of statistical inference which is called the maximum entropy estimate. It is the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information." Entropy-based inference is a form of Bayesian inference that is convenient when the data is sparse [2] and encapsulates common-sense reasoning [11]. It has three difficulties. First, it assumes that what the agent knows is "the sum total of the agent's knowledge, it is not a summary of the agent's knowledge, it is all there is" [11] — this assumption is referred to as *Watt's Assumption* [4]. Second, it may only be applied to a consistent set of beliefs. Third, its knowledge base is expressed in first-order logic, and so probability distributions are built over finite sample spaces. The way in which the sample space is chosen will affect the inferences drawn — this is referred to as *representation dependence* [3]. Despite these difficulties, maximum entropy inference is an elegant formulation of common sense reasoning that is useful when data is sparse.

3. VALUING INFORMATION

If β passes an utterance to α, α evaluates this act in two ways. First, it is valued for the strategic significance of the information that it contains, precisely it is measured as the expected increase in utility that α expects to enjoy given that it has the information — this is the *utilitarian measure*. Second, it is valued because the sending agent *was prepared to divulge* the information in the utterance, precisely it is measured as the decrease in uncertainty that the receiving agent has over the sending agent's private information — this is the *information measure*. All utterances received are qualified by α with a belief probability as described in Section 2.

From α's point of view, β's *private information* is everything that β knows and that α does not know with certainty. Due to the persistent effect of Equation 1, this will include much of what β knows.

[1]Sets of commitments between more than two agents are not considered here.

[2]One simple criterion is: $\mathbb{H}(\mathcal{M}^{t+1}) < \mathbb{H}(\mathcal{M}^t)$.

3.1 Utilitarian measures

An agent may wish to decide which action, $\{a_i\}$, to take where the payoff depends on which state, $\{s_j\}$, the world is in when the action is taken (possibly in the future). The payoff, $\vec{v_i}$, from taking action a_i is a vector where v_{ij} will be the payoff from taking action a_i and the state of the world is s_j. Let $\vec{p}$ be the probability mass function of a random variable representing the prior expectation about the state of the world when the action is taken. Then the *expected monetary value* gained by choosing action a_i is $m_i = \vec{p} \cdot \vec{v_i}$.

Armed with this information suppose that the agent applies some decision criterion, c, to decide what to do — perhaps c will choose the action with the greatest expected payoff: $\arg\max_i \vec{p} \cdot \vec{v_i}$. Now suppose that the agent receives information in an utterance u that enables him to refine his expectation of the state of the world when the action is to be taken $(\vec{p}|u)$, and that he applies the same criterion c. Then one *utilitarian value* of utterance u to criterion c is the difference between the payoffs of the respective outcomes. For each state of the world s_j let $b_j = \max_i v_{ij}$ i.e. b_j is the 'best' action that the agent can take if the state of the world is s_j then the *expected value of perfect information* is $\vec{p} \cdot \vec{b} - \max_i \vec{p} \cdot \vec{v_i}$; this is an upper limit on the total value of all possible utterances with respect to the application of criterion c.

Utilitarian measures of information are expressed in terms of: if you know information x when applying criterion y to determine which action to perform then you will gain utility z over not knowing x [8]. That is, they are defined in the context of some decision making act — they do not place an intrinsic value on information. So if an agent learns x at time t and is unaware of what future decisions he will make that will benefit from knowing x, then he will be unable to value x until he knows what those future decisions are. But, by the time he is aware of all of those decisions it may not be possible to reconstruct with certainty how he and the other agents would have behaved if he had *not* known x at time t. In summary, it is only possible to attach an intrinsic utilitarian value to information when the future decisions that are relevant to it are known.

We have described the value *gained* by acquiring information, we now consider the value *lost* by an agent's private information becoming public knowledge — that is, known to all agents in the system. Once information becomes public knowledge it has no tradable value until the integrity of the public's belief of it decays in time.

Utilitarian measures of information may be used when all the relevant future decisions are either known with certainty or a probability distribution expressing their likeliness to occur is known.

3.2 Information measures

α's world model, $\mathcal{M}^t$, is a set of probability distributions. If at time t, α receives an utterance u that may alter this world model (as described in Section 2) then the (Shannon) *information* in u with respect to the distributions in $\mathcal{M}^t$ is: $\mathbb{I}(u) = \mathbb{H}(\mathcal{M}^t) - \mathbb{H}(\mathcal{M}^{t+1})$. Let $\mathcal{N}^t \subseteq \mathcal{M}^t$ be α's model of agent β. If β sends the utterance u to α then the *information* about β within u is: $\mathbb{H}(\mathcal{N}^t) - \mathbb{H}(\mathcal{N}^{t+1})$. We give structure to the measurement of information using an *illocutionary framework* to categorise utterances, and an *ontology*.

The illocutionary framework will depend on the nature of the interactions between the agents. The LOGIC framework for argumentative negotiation [15] is based on five categories: Legitimacy of the arguments, Options i.e. deals that are acceptable, Goals i.e. motivation for the negotiation, Independence i.e: outside options, and Commitments that the agent has including its assets. The LOGIC framework contains two models per agent: first α's model of β's private information, and second, α's model of the private information that β has about α. Generally we assume that α has an illocutionary framework $\mathcal{F}$ and a categorising function $v : U \to \mathcal{P}(\mathcal{F})$ where U is the set of utterances. The power set, $\mathcal{P}(\mathcal{F})$, is required as some utterances belong to multiple categories. For example, in the LOGIC framework the utterance "I will not pay more for Protos[3] than the price that John charges" is categorised as both Option and Independence.

We assume an ontology, and $\mathcal{O}$ denotes its concepts that are organised in an *is-a* hierarchy.[4] δ measures the semantic distance between two concepts c_1 and c_2, for example [9]:

$$\delta(c_1, c_2) = e^{-\kappa_1 l} \cdot \frac{e^{\kappa_2 h} - e^{-\kappa_2 h}}{e^{\kappa_2 h} + e^{-\kappa_2 h}}$$

where l is the shortest path between the concepts, h is the depth of the deepest concept subsuming both concepts, and κ_1 and κ_2 are parameters scaling the contribution of shortest path length and depth respectively.

In [15] two central concepts are used to describe relationships and dialogues between a pair of agents. These are *intimacy* — degree of closeness, and *balance* — degree of fairness. Both of these concepts are summary measures of relationships and dialogues, and are expressed in the LOGIC framework as 5×2 matrices.

More generally, the intimacy of α's relationship with β_i, I_i^t, measures the amount that α knows about β_i's private information and is represented as real numeric values over $\mathcal{G} = \mathcal{F} \times \mathcal{O}$. Suppose α receives utterance u from β_i and that category $f \in v(u)$. For any concept $c \in \mathcal{O}$, define $\Delta(u, c) = \max_{c' \in u} \delta(c', c)$. Denote the value of I_i^t in position (f, c) by $I_{i(f,c)}^t$ then: $I_{i(f,c)}^t = \rho \times I_{i(f,c)}^{t-1} + (1 - \rho) \times \mathbb{I}(u) \times \Delta(u, c)$ for any c, where ρ is the discount rate. The *balance* of α's relationship with β_i, B_i^t, is the element by element numeric difference of I_i^t and α's estimate of β_i's intimacy on α.

[13] describes measures of: *trust* (in the execution of contracts), *honour* (validity of argumentation), and *reliability* (of information). The execution of contracts, soundness of argumentation and correctness of information are all represented as conditional probabilities $\mathbb{P}(\varphi'|\varphi)$ where φ is an expectation of what may occur, and φ' is the subsequent observation of what does occur.

[14] describes a single computational framework for these three measures that summarise α's observations of β's behaviour. One of these summary measures is:

$$M(\alpha, \beta, \varphi) = 1 - \sum_{\varphi'} \mathbb{P}_I^t(\varphi'|\varphi, e) \log \frac{\mathbb{P}_I^t(\varphi'|\varphi, e)}{\mathbb{P}^t(\varphi'|\varphi)}$$

where the "1" is an arbitrarily chosen constant being the maximum value that this measure may have, and $\mathbb{P}_I^t(\varphi'|\varphi, e)$ is a distribution of enactments that represent α's "ideal" in the sense that it is the best that α could reasonably expect to happen in *context* e. If α repeatedly observes φ' then the

[3]A fine wine from the 'Ribera del Duero' region, Spain.

[4]A simplified way of understanding an utterance u is as a set of concepts in $\mathcal{O}$, that is $u = \{c_i \mid c_i \in \mathcal{O}\}$.

amount of information that those observations convey about the associated commitments, φ, is the *mutual information*: $\mathbb{I}(\varphi';\varphi) = \mathbb{H}(\varphi') - \mathbb{H}(\varphi'|\varphi)$, this measures the mutual dependence of the two variables, where $\mathbb{I}(\varphi';\varphi) = \mathbb{I}(\varphi;\varphi')$.

These summary measures are all abstracted using the ontology; for example, "What is my trust of John for the supply of red wine?". These measures are also used to summarise the information in some of the categories in the illocutionary framework. For example, if these measures are used to summarise estimates $\mathbb{P}^t(\varphi'|\varphi)$ where φ is a deep motivation of β's (i.e. a Goal), or a summary of β's financial situation (i.e. a Commitment) then this contributes to a sense of trust at a deep social level.

3.3 Confidentiality

[15] advocates the controlled revelation of information as a way of managing the intensity of relationships. In Section 3.1 we noted that information that becomes public knowledge is worthless, and so respect of confidentiality is vital to maintaining the value of revealed private information. We have not yet described how to measure the extent to which one agent respects the confidentiality of another agent's information — that is, the strength of belief that another agent will respect the confidentially of my information: both by not passing it on, and by not using it so as to disadvantage me.

Consider the motivating example, α sells a case of Protos to β at cost, and asks β to treat the deal in confidence. Moments later another agent β' asks α to quote on a case of Protos — α might then reasonably increase his belief in the proposition that β had spoken to β'. Suppose further that α quotes β' a fair market price for the Protos and that β' rejects the offer — α may decide to further increase this belief. Moments later β offers to purchase another case of Protos for the same cost. α may then believe that β may have struck a deal with β' over the possibility of a cheap case of Protos.

Confidentiality is the mirror image of trust, honour and reliability that are all built by an agent "doing the right thing" — respect for confidentiality is built by an agent *not* doing the wrong thing. As human experience shows, validating respect for confidentiality is a tricky business. One proactive ploy is to start a false rumour (e.g. "My wife is a matador.") and to observe how it spreads. The following reactive approach builds on the Protos example above.

An agent will know when it passes confidential information to another, and it is reasonable to assume that the significance of the act of passing it on decreases in time. In this simple model we do not attempt to value the information passed as in Section 3. We simply note the amount of confidential information passed and observe any indications of a breach of confidence.

If α sends utterance u to β "in confidence", then u is categorised as f as described in Section 3.2. C_i^t measures the amount of confidential information that α passes to β_i in a similar way to the intimacy measure I_i^t described in Section 3.2: $C^t_{i(f,c)} = \rho \times C^{t-1}_{i(f,c)} + (1-\rho) \times \Delta(u,c)$, for any c where ρ is the discount rate; if no information is passed at time t then: $C^t_{i(f,c)} = \rho \times C^{t-1}_{i(f,c)}$. C_i^t represents the time-discounted amount of confidential information passed in the various categories.

α constructs a companion framework to C_i^t, L_i^t is as estimate of the amount of information leaked by β_i represented in $\mathcal{G}$. Having confided u in β_i, α designs update functions J_u^L for the L_i^t. In the absence of evidence imported by the J_u^L functions, each value in L_i^t decays by: $L^t_{i(f,c)} = \xi \times L^{t-1}_{i(f,c)}$, where ξ is in $[0,1]$ and probably close to 1. The J_u^L functions scan every observable utterance, u', from each agent β' for evidence of leaking the information u, $J_u^L(u') = \mathbb{P}(\beta' \text{ knows } u \mid u' \text{ is observed})$. As previously: $L^t_{i(f,c)} = \xi \times L^{t-1}_{i(f,c)} + (1-\xi) \times J_u^L(u') \times \Delta(u,c)$ for any c.

This simple model estimates C_i^t the amount of confidential information passed, and L_i^t the amount of presumed leaked, confidential information represented over $\mathcal{G}$. As with most things that information-based agents do, the 'magic' is in the specification of the J_u^L functions. A more exotic model would estimate "who trusts who more than who with what information" — this is what we have elsewhere referred to as a *trust network*. The feasibility of modelling a trust network depends substantially on how much detail each agent can observe in the interactions between other agents.

4. DELIBERATIVE MECHANISM

4.1 Plans

A *plan* p is $p(a_p, s_p, t_p, u_p, c_p, g_p)$ where:

- a_p is a conditional action sequence — i.e. it is conditional on future states of the world, and on the future actions of other agents. We think of plans as probabilistic statecharts in the normal way where the arcs from a state are labelled with "event / condition / action" leading into a Ⓟ symbol that represents the lottery, s_p, that determines the next state as described following:
- $s_p : S \to \mathbb{P}(S_p = s) \equiv \vec{s}$ where S is the set of states and S_p is a random variable denoting the state of the world when a_p terminates[5].
- $t_p : S \to \mathbb{P}(T_p = t) \equiv \vec{t}$ where T_p is a random variable denoting the time that a_p takes to execute and terminate for some finite set of positive time interval values for t.
- $u_p : S \to \mathbb{P}(U_p = u) \equiv \vec{u}$ where U_p is a random variable denoting the gross utility gain, excluding the cost of the execution of a_p for some finite set of utility values for u.
- $c_p : S \to \mathbb{P}(C_p = c) \equiv \vec{c}$ where C_p is a random variable denoting the cost of the execution of a_p for some finite set of cost values for c.
- $g_p : S \to \mathbb{P}(G_p = g) \equiv \vec{g}$ where G_p is a random variable denoting the expected information gain to α *and* to β of the dialogue that takes place during the execution of the plan each expressed in $\mathcal{G} = \mathcal{F} \times \mathcal{O}$.

The distributions above are estimated by observing the performance of the plans as we now describe.[6] In the absence of any observations the probability mass functions for S_p,

[5]For convenience we assume that all action sequences have a "time out" and so will halt after some finite time.

[6]An obvious simplification would be to use point estimates for t_p, u_p, c_p and each element of g_p, but that is too weak a model to enable comparison.

T_p, U_p, C_p and G_p all decay at each and every time step by Equation 1.

The implementation of a_p does not concern us. We do assume that the way in which the plans are implemented enables the identification of common algorithms and maybe common methods within different plans. Given two plans p and q, the function $\text{Sim}(p, q) \in [0, 1]$ measures the similarity of their action sequences a_p and a_q in the sense their performance parameters are expected to be correlated to some degree.

Estimating S_p. Denote the prior estimate by $\vec{s^t}$. When a plan terminates, or is terminated, the world will be in one of p's end states. Call that state z. Then the observed distribution for $\vec{s^{t+\delta t}}$ will have the value 1 in position z. On the basis of this observation the agent may be inclined to fix its estimate for s_z^{t+1} at γ where $s_z^t \leq \gamma \leq 1$. The posterior distribution $\vec{s^{t+1}}$ is defined as the distribution with minimum relative entropy with respect to $\vec{s^t}$: $\vec{s^{t+1}} = \arg\min_{\vec{r}} \sum_j r_j \log \frac{r_j}{s^t}$ that satisfies the constraint $s_z^{t+1} = \gamma$. If $\gamma = s_z^t$ then the posterior is the same as the prior. If $\gamma = 1$ then the posterior is certain with $\mathbb{H}(\vec{s^{t+1}}) = 0$. One neat way to calibrate γ is in terms of the resulting information gain; that is to measure γ in terms of the resulting *learning rate* μ:

$$\mathbb{H}(\vec{s^{t+1}}) = (1 - \mu) \times \mathbb{H}(\vec{s^t}) \tag{2}$$

where μ: $0 < \mu < 1$.

Estimating T_p, U_p, C_p and G_p. Just as for estimating S_p, when the plan terminates α will have observations for the values of these variables, and as a result may wish to increase the corresponding frequency in the posterior to some new value. Using the method described above for estimating S_p, the posterior distribution is the distribution with minimum relative entropy with respect to the prior subject to the constraint that the frequency corresponding to the observation is increased accordingly.

Further, for these four variables we use the $\text{Sim}(\cdot, \cdot)$ function to revise the estimates for 'nearby' plans. In [14] two methods for using a $\text{Sim}(\cdot, \cdot)$ function to revise estimates are described — the situation here is rather simpler. Consider the variable C_p. Applying the method in the paragraph 'Estimating S_p.', suppose a value had been observed for C_p and as a result of which c_j^{t+1} had been constrained to be γ. Consider any plan q for which $\text{Sim}(p, q) > 0$. Denote $\mathbb{P}(C_q = c)$ by $\vec{d}$. The posterior distribution $\vec{d^{t+1}}$ is defined as the distribution with minimum relative entropy with respect to $\vec{d^t}$: $\vec{d^{t+1}} = \arg\min_{\vec{r}} \sum_j r_j \log \frac{r_j}{d^t}$ that satisfies the constraint: $d_j^{t+1} = \gamma'$ where γ' is such that:

$$\mathbb{H}(\vec{d^{t+1}}) = (1 - \mu \times \text{Sim}(p, q)) \times \mathbb{H}(\vec{d^t}) \tag{3}$$

where $0 \leq \text{Sim}(p, q) \leq 1$ with higher values indicating greater similarity.

4.2 Planning

If an agent's needs could potentially be satisfied by more than one plan then a mechanism is required to select which plan to use. As the execution of plans incurs a cost we assume that α won't simply fire off every plan that may prove to be useful. A random variable, V_p, derived from the expectations of S_p, T_p, U_p, C_p, G_p and other estimates in $\mathcal{M}^t$ represents the agent's expectations of each plan's overall *performance*. V_p is expressed over some finite, numerical valuation space with higher values being preferred.

The mechanisms that we describe all operate by selecting plans stochastically. We assume that there is a set of P candidate plans $\{p_i\}$ with corresponding random variables V_{p_i} representing performance, and plan p_j is chosen with probability q_j where $\sum_k q_k = 1$. Let $\mathcal{N}^t = \{V_{p_k}^t\}_{k=1}^P$. The integrity of the performance estimates for random variable V_{p_i} are maintained using the method "Estimating S_p" in Section 4.1. If p_i is selected at time t then when it terminates the observed performance, $v_{p_i,\text{ob}}^t$, is fed into that method.

First, consider the naïve mechanism that selects plan p_j by: $q_j = 1$ for $j = \arg\max_i \mathbb{E}(V_{p_i})$. This mechanism is well-suited to a one-off situation. But if the agent has continuing need of a set of plans then choosing the plan with highest expected payoff may mean that some plans will not be selected for a while by which time their performance estimates will have decayed by Equation 1 to such a extent that may never be chosen. An agent faces the following dilemma: the only way to preserve a reasonably accurate estimate of plans is to select them sufficiently often — even if they they don't perform well today perhaps one day they will shine.

The simple method: $q_i = \frac{1}{P}$ selects all plans with equal probability. The following method attempts to prevent the uncertainty of estimates from decaying above a threshold, τ, by setting $q_j = 1$ where:
if $\exists i \cdot \mathbb{H}(V_{p_i}) > \tau$ **then** let $j = \arg\max_k \mathbb{H}(V_{p_k})$
else let $j = \arg\max_k \mathbb{E}(V_{p_k})$
this method may deliver poor performance from the '**then**' and good performance from the '**else**', but at least it attempts to maintain some level of integrity of the performance estimates, even if it does so in an elementary way.

A strategy is reported in [6] on how to place all of one's wealth as win-bets indefinitely on successive horse races so as to maximise the rate of growth; this is achieved by proportional gambling, i.e. by betting a proportion of one's wealth on each horse equal to the probability that that horse will win. This result is interesting as the strategy is independent of the betting odds. Whether it will make money will depend on the punter's ability to estimate the probabilities better than the bookmaker. The situation that we have is not equivalent to the horse race, but it is tempting to suggest the strategies:

$$q_i = \frac{\mathbb{E}(V_{p_i})}{\sum_k \mathbb{E}(V_{p_k})} \tag{4}$$

$$q_i = \mathbb{P}(V_{p_i} > V_{p_j}), \forall V_{p_j} \in \mathcal{N}, j \neq i \tag{5}$$

For the second strategy: q_i is the probability that p_i's performance is the better than that of all the other plans. With this definition it is clear that $\sum_i q_i = 1$. Both strategies will favour those plans with a better performance history. Whether they will prevent the integrity of the estimates for plans with a poor history from decaying to a meaningless level will depend on the value of λ in Equation 1, the value of μ in Equation 2, and on the frequency with which plans are activated. As the estimates for plans that perform well, and plans that perform badly, all decay to the maximum entropy decay limit $\mathbb{D}(V_{p_i})$ if they are not invoked, both of these strategies indirectly take account of the level of certainty in the various performance estimates.

We consider now the stability of the integrity of the performance estimates in time. If plan p_j is *not* executed the information loss in X_j^t for one time step due to the effect of Equation 1 is: $\lambda \times \mathbb{H}(X_j^t)$. If no plans in $\mathcal{N}$ are executed during one time step then the total information loss in $\mathcal{N}$ is: $\lambda \times \sum_k \mathbb{H}(X_k^t)$. If plan p_j *is* executed the information gain in X_j^t due to the effect of Equation 2 is: $\mu \times \mathbb{H}(X_j^t)$, but this observation may affect the other variables in $\mathcal{N}^t$ due to Equation 3, and the total information gain in $\mathcal{N}$ is: $\mu \times \sum_k \mathrm{Sim}(p_j, p_k) \times \mathbb{H}(X_k^t)$. Assuming that at most one plan in $\mathcal{N}^t$ is executed during any time step, and that the probability of one plan being executed in any time step is χ; the expected net information gain of $\mathcal{N}^{t+1}$ compared with $\mathcal{N}^t$ is:

$$\chi \cdot \mu \cdot \sum_j q_j \cdot \sum_k \mathrm{Sim}(p_j, p_k) \cdot \mathbb{H}(X_k^t) - \lambda \cdot \sum_k \mathbb{H}(X_k^t) \quad (6)$$

If this quantity is negative then the agent may decide to take additional steps to gain performance measurements so as to avoid the integrity of these estimates from consistently declining.

We now consider the parameters λ and μ to be used with the strategy in Equation 4. The effect of Equation 1 on variable V_i after t units of time is:

$$\left(1 - (1-\lambda)^t\right) \times \mathbb{D}(V_{p_i}) + (1-\lambda)^t \times V_{p_i}^{t_0}$$

The probability that plan p_i will be activated at any particular time is:

$$\chi \times \frac{\mathbb{E}(V_{p_i})}{\sum_k \mathbb{E}(V_{p_k})}$$

and the mean of these probabilities for all plans is: $\frac{\chi}{P}$. So the mean number of time units between each plan's activation is: $\frac{N}{\chi}$. In the absence of any intuitive value for λ, a convenient way to calibrate λ is in terms of the expected total decay towards $\mathbb{D}(V_{p_i})$ between each activation — this is expressed as some constant ϕ, where $0 < \phi < 1$. For example, $\phi = \frac{1}{2}$ means that we expect a 50% decay between activations. The value of λ that will achieve this is: $\lambda = 1 - (1-\phi)^{\chi \div N}$. Then the value for μ is chosen so that the expression (6) is non-negative. Using these values should ensure that the probability distributions for the random variables V_i remain within reasonable bounds, and so remain reasonably discriminating.

It would be nice to derive a method that was optimal in some sense, but this is unrealistic if the only data available is historic data such as the V_{p_i}. In real situations the past may predict the future to some degree, but can not be expected to predict performance outcomes that are a result of interactions with other autonomous agents in a changing environment. As a compromise, we propose to use (5) with values for λ and μ determined as above. (5) works with the whole distribution rather than (4) that works only with point estimates, but is algebraically simpler. These methods are proposed on the basis that the historic observations are all that α has.

5. STRATEGIES

An information-based agent's deliberative logic consists of:

1. The agent's *raison d'être* — its mission — this may not be represented in the agent's code, and may be implicit in the agent's design.

2. A set of *values*, Π, — high level principles — and a fuzzy function $\upsilon : (S \times A \times \Pi) \to$ fuz, that estimates, when the world is in state $s \in S$, whether the agent performing action $a \in A$ supports or violates a value $\pi \in \Pi$.

3. A *strategy* that provides an overarching context within which the plans are executed — see Section 5.2. The strategy is responsible for the evolution of the relationships between the agents, and for ensuring that plans take account of the state of those relationships.

4. A hierarchy[7] of *needs*, N, and a function $\sigma : N \to \mathcal{P}(S)$ were $\sigma(n)$ is the set of states that satisfy need $n \in N$. Needs turn 'on' spontaneously, and in response to *triggers*, T; they turn 'off' because the agent believes they are satisfied.

5. A set of *plans*, P — Section 4.1.

In this model an agent knows with certainty those states that will satisfy a need, but does *not* know with certainty what state the world is in. Before describing information-based strategies in Section 5.2 we discuss the role of preferences in managing the information revelation and discovery process.

5.1 The role of preferences

Agent α's *preferences* is a relation defined over an *outcome space*, where $s_1 \prec_\alpha s_2$ denotes "α prefers s_2 to s_1". Elements in the outcome space may be described either by the world being in a certain state or by a concept in the ontology having a certain value. If an agent knows its preferences then it may use results from game theory or decision theory to achieve a preferred outcome in some sense. For example, an agent may prefer the concept of price (from the ontology) to have lower values than higher, or to purchase wine when it is advertised at a discount (a world state). In practice the articulation of a preference relation may not be simple.

Consider the problem of specifying a preference relation for a collection of fifty cameras with different features, from different makers, with different prices, both new and second hand. This is a multi-issue evaluation problem. It is realistic to suggest that "a normal intelligent human being" may not be able to place the fifty cameras in a preference ordering with certainty, or even to construct a meaningful probability distribution to describe it. The complexity of articulating preferences over real negotiation spaces poses a practical limitation on the application of preference-based strategies.

In contract negotiation the outcome of the negotiation, (a', b'), is the enactment of the commitments, (a, b), in that contract, where a is α's commitment and b is β's. Some of the great disasters in market design [7], for example the Australian Foxtel fiasco, could have been avoided if the designers had considered how the agents were expected to deviate (a', b') from their commitments (a, b) after the contract is signed.

Consider a contract (a, b), and let $(\mathbb{P}_\alpha^t(a'|a), \mathbb{P}_\alpha^t(b'|b))$ denote α's estimate of what will be enacted if (a, b) is signed. Further assume that the pair of distributions $\mathbb{P}_\alpha^t(a'|a)$ and $\mathbb{P}_\alpha^t(b'|b)$ are independent[8] and that α is able to estimate

[7]In the sense of the well-known Maslow hierarchy, where the satisfaction of needs that are lower in the hierarchy take precedence over the satisfaction of needs that are higher.

[8]That is we assume that while α is executing commitment a she is oblivious to how β is executing commitment b and *vice versa*.

$\mathbb{P}^t_\alpha(a'|a)$ with confidence. α will only be confident in her estimate of $\mathbb{P}^t_\alpha(b'|b)$ if β's actions are constrained by norms, or if α has established a high degree of trust in β. If α is unable to estimate $\mathbb{P}^t_\alpha(b'|b)$ with reasonable certainty then put simply: she won't know what she is signing. For a utilitarian α, $(a_1, b_1) \prec_\alpha (a_2, b_2)$ if she prefers $(\mathbb{P}^t_\alpha(a'_2|a_2), \mathbb{P}^t_\alpha(b'_2|b_2))$ to $(\mathbb{P}^t_\alpha(a'_1|a_1), \mathbb{P}^t_\alpha(b'_1|b_1))$ in some sense.

One way to manage contract acceptance when the agent's preferences are unknown is to found the acceptance criterion instead on the simpler question: "how certain am I that (a, b) is a good contract to sign?" — under realistic conditions this is easy to estimate[9].

So far we have not considered the management of information exchange. When a negotiation terminates it is normal for agents to review what the negotiation has cost *ex post*; for example, "I got him to sign up, but had to tell him about our plans to close our office in Girona". In Section 3.1 we argued that it is not feasible to attach an intrinsic value to information that is related to the value derived from enactments. Without knowing what use the recipient will make of the "Girona information", it is not possible to relate the value of this act of information revelation to outcomes and so to preferences.

While this negotiation is taking place how is the agent to decide whether to reveal the "Girona information"? He won't know then whether the negotiation will terminate with a signed contract, or what use the recipient may be able to make of the information in future, or how any such use might affect him. In general it is unfeasible to form an expectation over these things. So we argue that the decision of whether to reveal a piece of information should *not* be founded on anticipated negotiation outcomes, and so this decision should not be seen in relation to the agent's preferences. The difficulty here is that value is derived from information in a fundamentally different way to the realisation of value from owning a commodity, for example[10].

An agent should reveal information if: it assists the dialogue towards a satisfactory conclusion, it deepens the relationship (if that is desired), he can confide in the recipient not to broadcast it, and can trust the recipient not to be malicious with it. In general, information revelation and discovery should be seen in the context of the evolution of the social relationship between the agents; it should *not* be managed by a utilitarian strategy.

A preference-based strategy may call upon powerful ideas from game theory. For example, to consider equilibria α will require estimates of $\mathbb{P}^t_\beta(a'|a)$ and $\mathbb{P}^t_\beta(b'|b)$ in addition to $\mathbb{P}^t_\alpha(a'|a)$ and $\mathbb{P}^t_\alpha(b'|b)$ — these estimates may well be even more speculative than those in the previous paragraph. In addition she will require knowledge about β's utility function. In simple situations this information may be known, but in general it will not.

5.2 Information-based strategies

We now describe the strategic reasoning of an information-based agent. This takes account of the, sometimes conflicting, utilitarian and information measures of utterances in dialogues and relationships. This general definition may be instantiated by specifying functions for the ψ_i in the following.

The following notation is used below. R^t_i denotes the relationship (i.e. the set of all dialogues) between α and β_i at time t. *Intimacy* is a summary measure of a relationship or a dialogue and is represented in $\mathcal{G}$. We write I^t_i to denote the intimacy of that relationship, and $I(d)$ to denote the intimacy of dialogue d. Likewise B^t_i and $B(d)$ denotes balance.

The Needs Model. α is driven by its needs. When a need fires, a plan is chosen to satisfy that need using the method in Section 4.2. If α is to contemplate the future she will need some idea of her future needs — this is represented in her *needs model*: $\nu : \mathcal{T} \to \times^n[0,1]$ where $\mathcal{T}$ is time, and: $\nu(t) = (n^t_1, \dots, n^t_N)$ where $n^t_i = \mathbb{P}(\text{need } i \text{ fires at time } t)$.

Setting Relationship Targets. On completion of each dialogue of which α is a part, she revises her aspirations concerning her intimacy with all the other agents. These aspirations are represented as a *relationship target*, T^t_i, for each β_i, that is represented in $\mathcal{G}$. Let $\mathbf{I}^t = (I^t_1, \dots, I^t_o)$, $\mathbf{B}^t = (B^t_1, \dots, B^t_o)$ and $\mathbf{T}^t = (T^t_1, \dots, T^t_o)$, then $\mathbf{T}^t = \psi_1(\nu, \mathbf{I}^t, \mathbf{B}^t)$ — this function takes account of all β_i and aims to encapsulate an answer to the question: "Given the state of my relationships with my trading partners, what is a realistic set of relationships to aim for in satisfaction of my needs?".

Activating Plans. If at time t, some of α's active needs, N^t_{active}, are not adequately[11] being catered for, N^t_{neglect}, by existing active plans, P^t_{active}, then select P^{t+1}_{active} to take account of those needs:

$$P^{t+1}_{\text{active}} = \psi_2(P^t_{\text{active}}, N^t_{\text{neglect}}, N^t_{\text{active}}, \mathbf{I}^t, \mathbf{T}^t)$$

The idea being that α will wish select P^{t+1}_{active} so as to move each observed intimacy I^t_i towards its relationship target intimacy T^t_i. Having selected a plan p, $\mathbb{E}(U_p)$ and $\mathbb{E}(G_p)$ assist α to set the *dialogue target*, D^t_i, for the current dialogue [15]. In Section 4.2 we based the plan selection process on a random variable V_p that estimates the plan's performance in some way. If α is preference-aware then V_p may be defined in terms of its preferences.

Deactivating Plans. If at time t, a subset of α's active plans, $P^t_{\text{sub}} \subset P^t_{\text{active}}$, adequately caters for α's active needs, N^t_{active}, then:

$$P^{t+1}_{\text{active}} = \psi_3(P^t_{\text{active}}, N^t_{\text{active}}, \mathbf{I}^t, \mathbf{T}^t)$$

is a minimal set of plans that adequately cater for N^t_{active} in

[9]In multi-issue negotiation an agent's preferences over each individual issue may be known with certainty. Eg: she may prefer to pay less than pay more, she may prefer to have some feature to not having it. In such a case, if some deals are known to be unacceptable with certainty, some are known to be acceptable with certainty, and, perhaps some known to be acceptable to some degree of certainty then maximum entropy logic may be applied to construct a complete distribution representing 'certainty of acceptability' over the complete deal space. This unique distribution will be *consistent* with what *is* known, and *maximally noncommittal* with respect to what is *not* known.

[10]If a dialogue is not concerned with the exchange of anything with utilitarian value, then the two agents may feel comfortable to balance the information exchanged using the methods in Section 3.2.

[11]For each need n, $\sigma(n)$ is the set of states that will satisfy n. For each active plan p, $\mathbb{P}(S_p = s)$ is probability distribution over the possible terminal states for p. During p's execution this initial estimation of the terminal state is revised by taking account of the known terminal states of executed sub-plans and $\mathbb{P}(S_{p'} = s)$ for currently active sub-plans p' chosen by p to satisfy sub-goals. In this way we continually revise the probability that P^{t-1}_{active} will satisfy α's active needs.

the sense described above. The idea here is that P^{t+1}_{active} will be chosen to best move the observed intimacy I^t_i towards the relationship target intimacy T^t_i as in the previous paragraph.

The work so far describes the selection of plans. Once selected a plan will determine the actions that α makes where an action is to transmit an utterance to some agent determined by that plan. Plans may be bound by interaction protocols specified by the host institution.

Executing a Plan — Options. [15] distinguishes between a *strategy* that determines an agent's Options from which a single kernel action, a, is selected; and *tactics* that wrap that action in argumentation, a^+ — that distinction is retained below. Suppose that α has adopted plan p that aims to satisfy need n, and that a dialogue d has commenced, and that α wishes to transmit some utterance, u, to some agent β_i. In a multi-issue negotiation, a plan p will, in general, determine a set of Options, $A^t_p(d)$ — if α is preference aware [Section 5.1] then this set could be chosen so that these options have similar utility. Select a from $A^t_p(d)$ by:

$$a = \psi_4(A^t_p(d), \Pi, D^t_i, I(d), B(d))$$

that is the action selected from $A^t_p(d)$ will be determined by α's set of values, Π, and the contribution a makes to the development of intimacy.

If d is a bilateral, multi-issue negotiation we note four ways that information may be used to select a from $A^t_p(d)$. (1) α may select a so that it gives β_i similar information gain as β_i's previous utterance gave to α. (2) If a is to be the opening utterance in d then α should avoid making excessive information revelation due to ignorance of β_i's position and should say as little as possible. (3) If a requires some response (e.g. a may be an offer for β_i to accept or reject) then α may select a to give her greatest expected information gain about β_i's private information from that response, where the information gain is either measured overall or restricted to some area of interest in $\mathcal{M}^t$. (4) If a is in response to an utterance a' from β_i (such as an offer) then α may use entropy-based inference to estimate the probability that she should accept the terms in a' using nearby offers for which she knows their acceptability with certainty [14].

Executing a Plan — Tactics. The previous paragraph determined a kernel action, a. Tactics are concerned with wrapping that kernel action in argumentation, a^+. To achieve this we look beyond the current action to the role that the dialogue plays in the development of the relationship:

$$a^+ = \psi_5(a, , T^t_i, I^t_i, I(d), B^t_i, B(d))$$

In [15] *stance* is meant as random noise applied to the action sequence to prevent other agent's from decrypting α's plans. Stance is important to the argumentation process but is not discussed here.

6. CONCLUSION

In this paper we have presented a number of measures to value information including a new model of confidentiality. We have introduced a planning framework based on the kernel components of an information-based agent architecture (i.e. decay, semantic similarity, entropy and expectations). We have defined the notion of strategy as a control level over the needs, values, plans and world model of an agent. Finally, the paper overall offers a model of negotiating agents that integrates previous work on information-based agency and that overcomes some limitations of utility-based architectures (e.g. preference elicitation or valuing information).

7. 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] A. Golan, G. Judge, and D. Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* Financial Economics and Quantitative Analysis. John Wiley and Sons, Inc., 1996.

[3] J. Halpern. *Reasoning about Uncertainty.* MIT Press, 2003.

[4] M. Jaeger. Representation independence of nonmonotonic inference relations. In *Proceedings of KR'96*, pages 461–472. Morgan Kaufmann, 1996.

[5] E. Jaynes. Information theory and statistical mechanics: Part I. *Physical Review*, 106:620 – 630, 1957.

[6] J. J. Kelly. A new interpretation of information rate. *IEEE Transactions on Information Theory*, 2(3):185–189, September 1956.

[7] P. Klemperer. What really matters in auction design. *The Journal of Economic Perspectives*, 16(1):169–189, 2002.

[8] D. Lawrence. *The Economic Value of Information.* Springer-Verlag, 1999.

[9] Y. Li, Z. A. Bandar, and D. McLean. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):871 – 882, July / August 2003.

[10] D. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.

[11] J. Paris. Common sense and maximum entropy. *Synthese*, 117(1):75 – 93, 1999.

[12] C. Sierra and J. Debenham. An information-based model for trust. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. Singh, and M. Wooldridge, editors, *Proceedings Fourth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2005*, pages 497 – 504, Utrecht, The Netherlands, July 2005. ACM Press, New York.

[13] C. Sierra and J. Debenham. Trust and honour in information-based agency. In P. Stone and G. Weiss, editors, *Proceedings Fifth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2006*, pages 1225 – 1232, Hakodate, Japan, May 2006. ACM Press, New York.

[14] C. Sierra and J. Debenham. Information-based agency. In *Proceedings of Twentieth International Joint Conference on Artificial Intelligence IJCAI-07*, pages 1513–1518, Hyderabad, India, January 2007.

[15] C. Sierra and J. Debenham. The LOGIC Negotiation Model. In *Proceedings Sixth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2007*, Honolulu, Hawai'i, May 2007.