Merging Intelligent Agency and the Semantic Web

John Debenham[1] and Carles Sierra[2]

[1] University of Technology, Sydney, Australia debenham@it.uts.edu.au
[2] Institut d'Investigació en Intel·ligència Artificial - IIIA,
Spanish Scientific Research Council, CSIC
08193 Bellaterra, Catalonia, Spain sierra@iiia.csic.es

Abstract. The semantic web makes unique demands on agency. Such agents should: be built around an ontology and should take advantage of the relations in it, be based on a grounded approach to uncertainty, be able to deal naturally with the issue of semantic alignment, and deal with interaction in a way that is suited to the co-ordination of services. A new breed of 'information-based' intelligent agents [1] meets these demands. This form of agency is founded on ideas from information theory, and was inspired by the insight that interaction is an information revelation and discovery process. Ontologies are fundamental to these agent's reasoning that relies on semantic distance measures. They employ entropy-based inference, a form of Bayesian inference, to manage uncertainty that they represent using probability distributions. Semantic alignment is managed through a negotiation process during which the agent's uncertain beliefs are continually revised. The co-ordination of services is achieved by modelling interaction as time-constrained, resource-constrained processes — a proven application of agent technology. In addition, measures of trust, reputation, and reliability are unified in a single model.

1 Introduction

The Semantic Web is a *data* sharing effort, a colossal human effort to liberate data currently confined to each one's private space. The main social objective of the Semantic Web is to permit the best possible data retrieval from this potentially huge distributed collection of data repositories. This extension of the classical *document* sharing approach of the web offers a great potential for human users but puts a series of technical challenges to the data retrieval tools: (1) heterogeneity in the representation of data, (2) inconsistency of the data appearing in different sites, and (3) uncertainty on the values associated to properties.

We argue in this paper that agent technology based on information theory is a sound way of addressing these challenges and at the same time is a feasible engineering approach to build actual web applications: agents permit to incorporate proactive behaviour into web services, or to co-ordinate in flexible ways P2P or Grid nodes, while facilitating a personalised interaction with users. Moreover, agent technology permits to keep track of interactions and provenance, as well as to build world models that facilitate the interpretation of the information gathered to assess the behaviour of other agents in the network. In this way agent technology can be used, as we show here, to give a clear operational meaning to the elusive top layer of the Semantic Web tower, that is, to the concept of *trust* [2].

Information-based agency is grounded on information-based concepts [3]. The agent architecture admits a game-theoretical reading and an information-theoretical reading. This approach contrasts with previous work on interaction that did not take information exchange into account, but focused on the similarity of offers [4, 5], game theory [6], or first-order logic [7]. This preoccupation with information and its integrity, together with the fundamental role played by ontologies, is the basis for their affinity with the semantic web. We use the following notation: a multiagent system $\{\alpha, \beta_1, \ldots, \beta_o, \xi, \theta_1, \ldots, \theta_t\}$, contains an agent α that interacts with other agents, β_i, information sources wrapped as agents, θ_j, and an *institutional agent*, ξ, that represents the prevailing norms of behaviour that may include laws and rules [8].

We will describe a *communication language* $\mathcal{C}$ that incorporates the specification of an *ontology* and permits us both to structure the dialogues and to structure the processing of the information gathered by agents. Agents have an *internal language* $\mathcal{L}$ used to build a probabilistic *world model*. We understand agents as being built on top of two basic functionalities. First, a *proactive machinery*, that transforms *needs* into *goals* and these into *plans* composed of *actions*. Second, a reactive machinery, that uses the received messages to revise the world model by updating the probability distributions in it. Agents summarise their world models using a number of measures (e.g. trust, reputation, and reliability [9]) that can then be used to define strategies for "exchanging information" — in the sense developed here, this is the only thing that an agent can do. Each agent has its own ontology that, together with the ontology identified as the context for each incoming illocution, plays a fundamental role in the agent's operation.

We introduce the communication language and its attendant ontological machinery in Section 2, the agent architecture in Section 3, measures of trust, reliability and reputation based on the architecture in Section 4, the way in which these agents deal with semantic alignment is described in Section 5, in Section 6 the agents co-ordinate services by employing goal-driven process management technology and treating the coordination problem as a time-constraiined, resource-constrained processes, and finally Section 7 concludes.

2 Communication Language $\mathcal{C}$

The shape of the language that α uses to represent the information received and the content of its dialogues depends on two fundamental notions. First, when agents interact within an overarching institution they explicitly or implicitly accept the *norms* that will constrain their behaviour, and accept the established sanctions and penalties whenever norms are violated. Second, the dialogues in which α engages are built around two fundamental actions: (i) passing information, and (ii) exchanging proposals and contracts. A *contract* is any agreement between two agents such as to provide some service over the Web or access to data — formally, a contract $\delta = (a, b)$ between agents α and β is a pair where a and b represent the activities that agents α and β are respectively responsible for. *Contracts* signed by agents and *information* passed by agents, are similar to norms in the sense that they oblige agents to behave in a particular way, so as to satisfy the conditions of the contract, or to make the world consistent with the information passed. Contracts and Information can then be thought of as normative statements that restrict an agent's behaviour.

Norms, contracts, and information have an obvious temporal dimension. Thus, an agent has to abide by a norm while it is operating within the Semantic web, a contract has a validity period, and a piece of information is true only during an interval in time. The set of norms affecting the behaviour of an agent define the *context* that the agent has to take into account.

The communication language that α needs requires two fundamental primitives: $\mathrm{Commit}(\alpha, \beta, \varphi)$ to represent, in φ, the state of affairs that α aims to bring about and that β has the right to verify, complain about or claim compensation for any deviations from, and $\mathrm{Done}(a)$ to represent the event that a certain action a[3] has taken place. In this way, norms, contracts, and information chunks will be represented as instances of $\mathrm{Commit}(\cdot)$ where α and β can be individual agents or the Semantic Web, $\mathcal{C}$ is:

$$\begin{aligned}
a &::= \mathit{illoc}(\alpha, \beta, \varphi, t) \mid a; a \mid \textbf{Let}\ \mathit{context}\ \textbf{In}\ a\ \textbf{End} \\
\varphi &::= \mathit{term} \mid \mathrm{Done}(a) \mid \mathrm{Commit}(\alpha, \beta, \varphi) \mid \varphi \wedge \varphi \mid \\
&\quad\ \ \varphi \vee \varphi \mid \neg\varphi \mid \forall v.\varphi_v \mid \exists v.\varphi_v \\
\mathit{context} &::= \varphi \mid \mathit{id} = \varphi \mid \mathit{prolog_clause} \mid \mathit{context}; \mathit{context}
\end{aligned}$$

where φ_v is a formula with free variable v, illoc is any appropriate set of illocutionary particles, ';' means sequencing, and *context* represents either previous agreements, previous illocutions, or code that aligns the ontological differences between the speakers needed to interpret an action a.

For example, we can represent the following offer: "If you spend a total of more than €100 on my information service during October then I will give you a 10% discount on all of my services in November", as:

$$\begin{aligned}
&\text{Offer}(\ \alpha, \beta, \mathtt{spent}(\beta, \alpha, \mathtt{October}, \text{X}) \wedge \text{X} \geq €100 \rightarrow \\
&\qquad \forall\ \text{y. Done(Inform}(\xi, \alpha, \text{pay}(\beta, \alpha, \text{y}), \text{November})) \rightarrow \\
&\qquad\quad \text{Commit}(\alpha, \beta, \text{discount(y,10\%)}))
\end{aligned}$$

Note the use of the institution agent ξ to report the payment.

2.1 The Ontological Context

In order to define the language introduced above that structures agent dialogues we need an ontology that includes a (minimum) repertoire of elements: a set of *concepts* (e.g. quantity, quality, material) organised in a is-a hierarchy (e.g. platypus is a mammal, australian-dollar is a currency), and a set of relations over these concepts (e.g. price(beer,AUD)).[4] We model ontologies following an algebraic approach [10] as:

An ontology is a tuple $\mathcal{O} = (C, R, \leq, \sigma)$ where:

[3] Without loss of generality we will assume that all actions are dialogical.

[4] Usually, a set of axioms defined over the concepts and relations is also required. We will omit this here.

1. C is a finite set of concept symbols (including basic data types);
2. R is a finite set of relation symbols;
3. $\leq$ is a reflexive, transitive and anti-symmetric relation on C (a partial order)
4. $\sigma : R \to C^+$ is the function assigning to each relation symbol its arity

where $\leq$ is a traditional *is-a* hierarchy, and R contains relations between the concepts in the hierarchy.

The concepts within an agent's ontology are closer, semantically speaking, depending on how far away are they in the structure defined by the $\leq$ relation. Semantic distance plays a fundamental role in strategies for information-based agency. How signed contracts, *Commit*$(\cdot)$ about objects in a particular semantic region, and their execution *Done*$(\cdot)$, *affect* our decision making process about signing future contracts on nearby semantic regions is crucial to model the common sense that human beings apply in managing business relationships. A measure [11] bases the *semantic similarity* between two concepts on the path length induced by $\leq$ (more distance in the $\leq$ graph means less semantic similarity), and the *depth* of the subsumer concept (common ancestor) in the shortest path between the two concepts (the deeper in the hierarchy, the closer the meaning of the concepts). For agent α semantic similarity could then be defined as:

$$\mathrm{Sim}(c, c', \mathcal{O}_\alpha) = e^{-\kappa_1 l} \cdot \frac{e^{\kappa_2 h} - e^{-\kappa_2 h}}{e^{\kappa_2 h} + e^{-\kappa_2 h}}$$

where l is the length (i.e. number of hops) of the shortest path between the concepts in $\mathcal{O}_\alpha$, 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.

Given a formula $\varphi \in \mathcal{C}$ in the communication language and an ontology $\mathcal{O}_\alpha$ we define the vocabulary or *ontological context* of the formula, $C(\varphi, \mathcal{O}_\alpha)$, as the set of concepts in $\mathcal{O}_\alpha$ used in φ. Thus, we extend the previous definition of similarity to sets of concepts in the following way:

$$\mathrm{Sim}(\varphi, \psi, \mathcal{O}_\alpha) = \max_{c_i \in C(\varphi, \mathcal{O}_\alpha)} \min_{c_j \in C(\psi, \mathcal{O}_\alpha)} \{\mathrm{Sim}(c_i, c_j, \mathcal{O}_\alpha)\}$$

The following relies on *a* measure of semantic distance but not necessarily this one.

3 Agent Architecture

The vision here is intelligent agents managing each service across the Semantic Web. An agent α receives all messages expressed in $\mathcal{C}$ in an in-box $\mathcal{X}$ where they are time-stamped and sourced-stamped. A message μ from agent β (or θ or ξ), expressed in the sender's ontology, is then moved from $\mathcal{X}$ to a *percept repository* $\mathcal{Y}^t$ where it is appended with a subjective belief function $\mathbb{R}^t(\alpha, \beta, \mu)$ that normally decays with time. α acts in response to a message that expresses a *need*. A need may be exogenous such as a need for information and may be triggered by another agent offering to supply it, or endogenous such as α deciding that it wishes to offer its information or services across the Web. Needs trigger α's goal/plan proactive reasoning described in Section 3.1, other messages are dealt with by α's reactive reasoning described in Section 3.2.

Fig. 1. The *information-based* agent architecture in summary — the notation has been simplified from the complete version in the text.

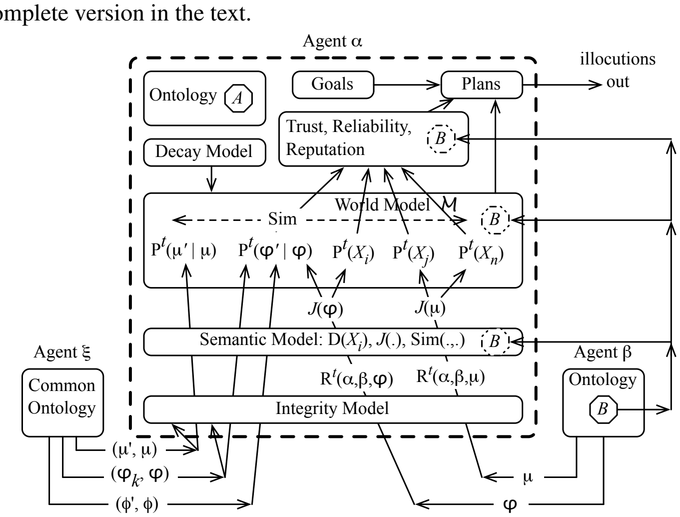

Figure 1 shows the agent architecture of agent α using ontology A in summary using notation that has been simplified from the version in the text. All communication is in the illocutionary communication language $\mathcal{C}$ — Section 2 — in the context of an ontology that is the foundation for the Semantic Model and the Sim similarity measure — Section 2.1. Agent β sends information μ to agent α in the context of ontology B, the integrity model adds a belief to it and the semantic model converts it to a set of constraints $\{J_s^{X_i}(\cdot)\}_{i\in T(s)}$ on the set of distributions $\{X_i\}_{i\in T(s)}$ in the world model — Section 3.2 — determined by each active plan s. Agent β makes commitment φ to agent α — agent α's expectation of what will occur is the distribution $\mathbb{P}^t(\varphi'|\varphi)$ — Section 3.4. The institution agent ξ reports on what occurs — (μ', μ) and (φ_k, φ) — that feeds into the Integrity Model — Section 3.3. Given β's commitment φ, the outcome, (φ_k, φ) updates α's expectation $\mathbb{P}^{t+1}(\varphi'|\varphi)$ using the method in Section 3.4 — likewise all of ξ's reports update other distributions in a way that is moderated by semantic distance as denoted by the double arrow labelled "Sim" that is derived from the ontology B. The decay model is described in Section 3.1. General measures of trust, reliability and reputation, that are so important on the Semantic Web, are given in Section 4 and summarise the World Model $\mathcal{M}^t$. α uses these summary measures and the world model to feed into its plans as described below.

3.1 Information Integrity

The Semantic Web provides a context of changing uncertainty and so α's goal/plan machinery typically will pursue multiple sub-goals concurrently. This applies both to securing agreements with service providers, and to interaction with public information sources that either may be unreliable or may take an unpredictable time to respond.

Each plan contains constructors for a *world model* $\mathcal{M}^t$ that consists of probability distributions, (X_i), in first-order probabilistic logic $\mathcal{L}$. $\mathcal{M}^t$ is then maintained from percepts received using *update functions* that transform percepts into constraints on $\mathcal{M}^t$ — described in Section 3.2.

The distributions in $\mathcal{M}^t$ are determined by α's plans that are determined by its needs. If α is negotiating some contract δ in satisfaction of need χ then it may require the distribution $\mathbb{P}^t(\text{eval}(\alpha, \beta, \chi, \delta) = e_i)$ where for a particular δ, $\text{eval}(\alpha, \beta, \chi, \delta)$ is an evaluation over some complete and disjoint *evaluation space* $E = (e_1, \ldots, e_n)$ that may contain hard (possibly utilitarian) values, or fuzzy values such as "reject" and "accept". This distribution assists α's strategies to decide whether to accept a proposed contract leading to a probability of acceptance $\mathbb{P}^t(\text{acc}(\alpha, \beta, \chi, \delta))$.

For example, $\mathbb{P}^t(\text{eval}(\alpha, \beta, \chi, \delta) = e_i)$ could be derived from the subjective estimate $\mathbb{P}^t(\text{satisfy}(\alpha, \beta, \chi, \delta) = f_j)$ of the expected extent to which the *execution* of δ by β will satisfy χ, and an objective estimate $\mathbb{P}^t(\text{val}(\alpha, \beta, \delta) = g_k)$ of the expected valuation of the *execution* of δ possibly in utilitarian terms. This second estimate could be derived by proactive reference to the $\{\theta_i\}$ for prevailing service pricing. In a negotiation α's plans may also construct the distribution $\mathbb{P}^t(\text{acc}(\beta, \alpha, \delta))$ that estimates the probability that β would accept δ — we show in Section 3.2 how α may derive this estimate from the information in β's proposals.

α's plans may construct various other distributions such as: $\mathbb{P}^t(\text{trade}(\alpha, \beta, o) = e_i)$ that β is a good agent to sign contracts with in context o, and $\mathbb{P}^t(\text{confide}(\alpha, \beta, o) = f_j)$ that α can trust β with confidential information the context o that consists of the interaction history and β's ontology $\mathcal{O}_\beta$.

The integrity of percepts decreases in time. α may have background knowledge concerning the expected integrity of a percept as $t \to \infty$. Such background knowledge will be expressed in terms of α's ontology $\mathcal{O}_\alpha$, and is represented as a *decay limit distribution*. If the background knowledge is incomplete then one possibility is for α to assume that the decay limit distribution has maximum entropy whilst being consistent with the data. Given a distribution, $\mathbb{P}(X_i)$, and a decay limit distribution $\mathbb{D}(X_i)$, $\mathbb{P}(X_i)$ decays by:

$$\mathbb{P}^{t+1}(X_i) = \Delta_i(\mathbb{D}(X_i), \mathbb{P}^t(X_i)) \tag{1}$$

where Δ_i is the *decay function* for the X_i satisfying the property that $\lim_{t\to\infty} \mathbb{P}^t(X_i) = \mathbb{D}(X_i)$. For example, Δ_i could be linear: $\mathbb{P}^{t+1}(X_i) = (1 - \nu_i) \times \mathbb{D}(X_i) + \nu_i \times \mathbb{P}^t(X_i)$, where $\nu_i < 1$ is the decay rate for the i'th distribution. Either the decay function or the decay limit distribution could also be a function of time: Δ_i^t and $\mathbb{D}^t(X_i)$.

3.2 New Information

In the absence of in-coming messages the integrity of $\mathcal{M}^t$ decays by Eqn. 1. The following procedure updates $\mathcal{M}^t$ for all percepts expressed in $\mathcal{C}$. Suppose that α receives a message μ from agent β in terms of ontology $\mathcal{O}_\beta$ at time t. Suppose that this message states that something is so with probability z, and suppose that α attaches an epistemic belief $\mathbb{R}^t(\alpha, \beta, \mu)$ to μ — this probability reflects α's level of personal *caution*. 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 $J_s^{X_i}(\mu)$ is a set of linear constraints on the posterior distribution for X_i. Examples of these update functions are given in Section 3.4. Denote the prior distribution $\mathbb{P}^t(X_i)$ by $\boldsymbol{p}$, and let $\boldsymbol{p}_{(\mu)}$ be the distribution with minimum relative entropy[5] with respect to $\boldsymbol{p}$: $\boldsymbol{p}_{(\mu)} = \arg\min_{\boldsymbol{r}} \sum_j r_j \log \frac{r_j}{p_j}$ that satisfies the constraints $J_s^{X_i}(\mu)$. Then let $\boldsymbol{q}_{(\mu)}$ be the distribution:

$$\boldsymbol{q}_{(\mu)} = \mathbb{R}^t(\alpha, \beta, \mu) \times \boldsymbol{p}_{(\mu)} + (1 - \mathbb{R}^t(\alpha, \beta, \mu)) \times \boldsymbol{p} \tag{2}$$

and then let:

$$\mathbb{P}^t(X_{i(\mu)}) = \begin{cases} \boldsymbol{q}_{(\mu)} & \text{if } \boldsymbol{q}_{(\mu)} \text{ is more interesting than } \boldsymbol{p} \\ \boldsymbol{p} & \text{otherwise} \end{cases} \tag{3}$$

A general measure of whether $\boldsymbol{q}_{(\mu)}$ is more interesting than $\boldsymbol{p}$ is: $\mathbb{K}(\boldsymbol{q}_{(\mu)} \| \mathbb{D}(X_i)) > \mathbb{K}(\boldsymbol{p} \| \mathbb{D}(X_i))$, where $\mathbb{K}(\boldsymbol{x} \| \boldsymbol{y}) = \sum_j x_j \ln \frac{x_j}{y_j}$ is the Kullback-Leibler distance between two probability distributions $\boldsymbol{x}$ and $\boldsymbol{y}$.

Finally merging Eqn. 3 and Eqn. 1 we obtain the method for updating a distribution X_i on receipt of a message μ:

$$\mathbb{P}^{t+1}(X_i) = \Delta_i(\mathbb{D}(X_i), \mathbb{P}^t(X_{i(\mu)})) \tag{4}$$

This procedure deals with integrity decay, and with two probabilities: first, the probability z in the percept μ, and second the belief $\mathbb{R}^t(\alpha, \beta, \mu)$ that α attached to μ.

In a simple multi-issue contract negotiation α may estimate $\mathbb{P}^t(\text{acc}(\beta, \alpha, \delta))$, the probability that β would accept δ, by observing β's responses. Using shorthand notation, if β sends the message $\text{Offer}(\delta_1)$ then α may derive the constraint:

$$J^{\text{acc}(\beta,\alpha,\delta)}(\text{Offer}(\delta_1)) = \{\mathbb{P}^t(\text{acc}(\beta, \alpha, \delta_1)) = 1\},$$

and if this is a counter offer to a former offer of α's, δ_0, then: $J^{\text{acc}(\beta,\alpha,\delta)}(\text{Offer}(\delta_1)) = \{\mathbb{P}^t(\text{acc}(\beta, \alpha, \delta_0)) = 0\}$. In the not-atypical special case of multi-issue contracting where the agents' preferences over the individual issues *only* are known and are complementary to each other's, maximum entropy reasoning can be applied to estimate the probability that any multi-issue δ will be acceptable to β by enumerating the possible worlds that represent β's "limit of acceptability" [14].

3.3 Reliability of an Information Source

$\mathbb{R}^t(\alpha, \beta, \mu)$ is an epistemic probability that represents α's belief in the validity of μ taking account of α's personal caution concerning β. An empirical estimate of $\mathbb{R}^t(\alpha, \beta, \mu)$

[5] Given a probability distribution $\boldsymbol{q}$, the *minimum relative entropy distribution* $\boldsymbol{p} = (p_1, \ldots, p_I)$ subject to a set of J linear constraints $\boldsymbol{g} = \{g_j(\boldsymbol{p}) = \boldsymbol{a_j} \cdot \boldsymbol{p} - c_j = 0\}, j = 1, \ldots, J$ (that must include the constraint $\sum_i p_i - 1 = 0$) is: $\boldsymbol{p} = \arg\min_{\boldsymbol{r}} \sum_j r_j \log \frac{r_j}{q_j}$. This may be calculated by introducing Lagrange multipliers $\boldsymbol{\lambda}$: $L(\boldsymbol{p}, \boldsymbol{\lambda}) = \sum_j p_j \log \frac{p_j}{q_j} + \boldsymbol{\lambda} \cdot \boldsymbol{g}$. Minimising L, $\{\frac{\partial L}{\partial \lambda_j} = g_j(\boldsymbol{p}) = 0\}, j = 1, \ldots, J$ is the set of given constraints $\boldsymbol{g}$, and a solution to $\frac{\partial L}{\partial p_i} = 0, i = 1, \ldots, I$ leads eventually to $\boldsymbol{p}$. Entropy-based inference is a form of Bayesian inference that is convenient when the data is sparse [12] and encapsulates common-sense reasoning [13].

may be obtained by measuring the 'difference' between commitment and enactment. Suppose that μ is received from agent β at time u and is verified by ξ as μ' at some later time t. Denote the prior $\mathbb{P}^u(X_i)$ by $\boldsymbol{p}$. Let $\boldsymbol{p}_{(\mu)}$ be the posterior minimum relative entropy distribution subject to the constraints $J_s^{X_i}(\mu)$, and let $\boldsymbol{p}_{(\mu')}$ be that distribution subject to $J_s^{X_i}(\mu')$. We now estimate what $\mathbb{R}^u(\alpha, \beta, \mu)$ should have been in the light of knowing *now*, at time t, that μ should have been μ'.

The idea of Eqn. 2, is that $\mathbb{R}^t(\alpha, \beta, \mu)$ should be such that, *on average* across $\mathcal{M}^t$, $\boldsymbol{q}_{(\mu)}$ will predict $\boldsymbol{p}_{(\mu')}$ — no matter whether or not μ was used to update the distribution for X_i, as determined by the condition in Eqn. 3 at time u. The *observed reliability* for μ and distribution X_i, $\mathbb{R}^t_{X_i}(\alpha, \beta, \mu)|\mu'$, on the basis of the verification of μ with μ', is the value of k that minimises the Kullback-Leibler distance:

$$\mathbb{R}^t_{X_i}(\alpha, \beta, \mu)|\mu' = \arg\min_k \mathbb{K}(k \cdot \boldsymbol{p}_{(\mu)} + (1-k) \cdot \boldsymbol{p} \parallel \boldsymbol{p}_{(\mu')})$$

The predicted *information* in the enactment of μ with respect to X_i is:

$$\mathbb{I}^t_{X_i}(\alpha, \beta, \mu) = \mathbb{H}^t(X_i) - \mathbb{H}^t(X_{i(\mu)}) \tag{5}$$

that is the reduction in uncertainty in X_i where $\mathbb{H}(\cdot)$ is Shannon entropy. Eqn. 5 takes account of the value of $\mathbb{R}^t(\alpha, \beta, \mu)$.

If $\mathbf{X}(\mu)$ is the set of distributions that μ affects, then the *observed reliability* of β on the basis of the verification of μ with μ' is:

$$\mathbb{R}^t(\alpha, \beta, \mu)|\mu' = \frac{1}{|\mathbf{X}(\mu)|} \sum_i \mathbb{R}^t_{X_i}(\alpha, \beta, \mu)|\mu' \tag{6}$$

If $\mathbf{X}(\mu)$ are independent the predicted *information* in μ is:

$$\mathbb{I}^t(\alpha, \beta, \mu) = \sum_{X_i \in \mathbf{X}(\mu)} \mathbb{I}^t_{X_i}(\alpha, \beta, \mu) \tag{7}$$

Suppose α sends message μ to β where μ is α's private information, then assuming that β's reasoning apparatus, but *not* β's ontology, mirrors α's, α can estimate $\mathbb{I}^t(\beta, \alpha, \mu)$.

For each formula φ at time t when μ has been verified with μ', the *observed reliability* that α has for agent β in φ is:

$$\mathbb{R}^{t+1}(\alpha, \beta, \varphi) = (1-\nu) \times \mathbb{R}^t(\alpha, \beta, \varphi) + \nu \times \mathbb{R}^t(\alpha, \beta, \mu)|\mu' \times \mathrm{Sim}(\varphi, \mu, \mathcal{O}_\beta)$$

where Sim measures the semantic distance between two sections of the ontology $\mathcal{O}_\beta$ as introduced in Section 2, and ν is the learning rate. Over time, α notes the context of the various μ received from β, and over the various contexts calculates the relative frequency, $\mathbb{P}^t(\mu)$. This leads to an overall expectation of the *reliability* that agent α has for agent β: $\mathbb{R}^t(\alpha, \beta) = \sum_\mu \mathbb{P}^t(\mu) \times \mathbb{R}^t(\alpha, \beta, \mu)$.

3.4 Expectation and Execution

The interaction between agents α and β will involve β making contractual commitments and (perhaps implicitly) committing to the truth of information exchanged. No matter what these commitments are, α will be interested in any variation between β's commitment, φ, and what is actually observed (as advised by the institution agent ξ), as the enactment, φ'. We denote the relationship between commitment and enactment, $\mathbb{P}^t(\text{Observe}(\varphi')|\text{Commit}(\varphi))$ simply as $\mathbb{P}^t(\varphi'|\varphi) \in \mathcal{M}^t$.

In the absence of in-coming messages the conditional probabilities, $\mathbb{P}^t(\varphi'|\varphi)$, should tend to ignorance as represented by the *decay limit distribution* and Eqn. 1. Eqn. 4 is used to revise $\mathbb{P}^t(\varphi'|\varphi)$ as observations are made. Let the set of possible enactments be $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_m\}$ with prior distribution $\boldsymbol{p} = \mathbb{P}^t(\varphi'|\varphi)$. Suppose that message μ is received, we estimate the posterior $\boldsymbol{p}_{(\mu)} = (p_{(\mu)i})_{i=1}^m = \mathbb{P}^{t+1}(\varphi'|\varphi)$.

First, if $\mu = (\varphi_k, \varphi)$ is observed then α may use this observation to estimate $p_{(\varphi_k)k}$ as some value d at time $t+1$. We estimate the distribution $\boldsymbol{p}_{(\varphi_k)}$ by applying the principle of minimum relative entropy as in Eqn. 4 with prior $\boldsymbol{p}$, and the posterior $\boldsymbol{p}_{(\varphi_k)} = (p_{(\varphi_k)j})_{j=1}^m$ satisfying the single constraint: $J^{(\varphi'|\varphi)}(\varphi_k) = \{p_{(\varphi_k)k} = d\}$.

Second, we consider the effect that the enactment ϕ' of another commitment ϕ, also by agent β, has on $\boldsymbol{p}$. This is achieved in two ways, first by appealing to the structure of the ontology using the $\text{Sim}(\cdot)$ function, and second by introducing a valuation function.

The $\text{Sim}(\cdot)$ *method.* Given the observation $\mu = (\phi', \phi)$, define the vector $\boldsymbol{t}$ by

$$t_i = \mathbb{P}^t(\varphi_i|\varphi) + (1- \mid \text{Sim}(\phi', \phi, \mathcal{O}_\beta) - \text{Sim}(\varphi_i, \varphi, \mathcal{O}_\beta) \mid) \cdot \text{Sim}(\varphi', \phi, \mathcal{O}_\beta)$$

for $i = 1, \dots, m$. $\boldsymbol{t}$ is not a probability distribution. The factor $\text{Sim}(\varphi', \phi, \mathcal{O}_\beta)$ limits the variation of probability to those formulae whose ontological context is not too far away from the observation. The posterior $\boldsymbol{p}_{(\phi',\phi)}$ is defined to be the normalisation of $\boldsymbol{t}$.

The valuation method. α may wish to value φ in some sense. This value will depend on the future use that α makes of it. So α estimates the value of φ using a probability distribution $(p_1, \dots, p_n)$ over some *evaluation space* $E = (e_1, \dots, e_n)$. $p_i = w_i(\varphi)$ is the probability that e_i is the correct evaluation of the enactment φ, and $\boldsymbol{w} : \mathcal{L} \times \mathcal{L} \to [0,1]^n$ is the *evaluation function*.

For a given φ_k, $(\mathbb{P}^t(\varphi_1|\varphi_k), \dots, \mathbb{P}^t(\varphi_m|\varphi_k))$ is the prior distribution of α's estimate of what will be observed if β committed to φ_k. $\boldsymbol{w}(\varphi_k) = (w_1(\varphi_k), \dots, w_n(\varphi_k))$ is α's evaluation over E of β's commitment φ_k. α's expected evaluation of what will be observed that β has committed to φ_k is $\boldsymbol{w}^{\text{exp}}(\varphi_k)$: $w^{\text{exp}}(\varphi_k)_i = \sum_{j=1}^m \mathbb{P}^t(\varphi_j|\varphi_k) \cdot w_i(\varphi_j)$ for $i = 1, \dots, n$. Now suppose that α observes the enactment ϕ' of another commitment ϕ also by agent β. Eg: α may acquire information about both the weather and the stock market from the same supplier. α may wish to revise the prior estimate of the expected valuation $\boldsymbol{w}^{\text{exp}}(\varphi_k)$ in the light of the observation (ϕ', ϕ) to:

$$\begin{aligned}&(\boldsymbol{w}^{\text{rev}}(\varphi_k) \mid (\phi'|\phi)) = \\ &\quad \boldsymbol{g}(\boldsymbol{w}^{\text{exp}}(\varphi_k), \text{Sim}(\phi', \phi, \mathcal{O}_\beta), \text{Sim}(\varphi, \phi, \mathcal{O}_\beta), \boldsymbol{w}(\varphi), \boldsymbol{w}(\phi), \boldsymbol{w}(\phi'))\end{aligned}$$

for some function $\boldsymbol{g}$ — the idea being, for example, that if the commitment, ϕ, to supply accurate weather information was not kept by β then α's expectation that the commitment, φ, to supply accurate stock market information should decrease. We estimate the posterior $\boldsymbol{p}_{(\phi',\phi)}$ by applying the principle of minimum relative entropy as in Eqn. 4 with prior $\boldsymbol{p}$ and $\boldsymbol{p}_{(\phi',\phi)} = (p_{(\phi',\phi)j})_{j=1}^{m}$ satisfying the n constraints:

$$J^{(\varphi'|\varphi)}((\phi',\phi)) = \Big\{ \sum_{j=1}^{m} p_{(\phi',\phi)j} \cdot w_i(\varphi_j) = \\ g_i(\boldsymbol{w}^{\exp}(\varphi_k), \mathrm{Sim}(\phi',\phi,\mathcal{O}_\beta), \mathrm{Sim}(\varphi,\phi,\mathcal{O}_\beta), \boldsymbol{w}(\varphi), \boldsymbol{w}(\phi), \boldsymbol{w}(\phi')) \Big\}_{i=1}^{n}$$

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.

4 Trust, Reliability and Reputation

The measures here generalise what are commonly called *trust*, *reliability* and *reputation* measures into a single computational framework that may be applied to information, information sources and information suppliers across the Semantic Web. It they are applied to the execution of contracts they become trust measures, to the validation of information they become reliability measures, and to socially transmitted overall behaviour they become reputation measures.

Ideal enactments. Consider a distribution of enactments that represent α's "ideal" in the sense that it is the best that α could reasonably expect to happen. This distribution will be a function of α's *context* with β denoted by e, and is $\mathbb{P}^t_I(\varphi'|\varphi, e)$. Here we measure the relative entropy between this ideal distribution, $\mathbb{P}^t_I(\varphi'|\varphi, e)$, and the distribution of expected enactments, $\mathbb{P}^t(\varphi'|\varphi)$. That is:

$$M(\alpha,\beta,\varphi) = 1 - \sum_{\varphi'} \mathbb{P}^t_I(\varphi'|\varphi,e) \log \frac{\mathbb{P}^t_I(\varphi'|\varphi,e)}{\mathbb{P}^t(\varphi'|\varphi)} \tag{8}$$

where the "1" is an arbitrarily chosen constant being the maximum value that this measure may have. This equation measures one, single commitment φ. It makes sense to aggregate these values over a class of commitments, say over those φ that are in the context o, that is $\varphi \le o$:

$$M(\alpha,\beta,o) = 1 - \frac{\sum_{\varphi:\varphi\le o} \mathbb{P}^t_\beta(\varphi)\,[1 - M(\alpha,\beta,\varphi)]}{\sum_{\varphi:\varphi\le o} \mathbb{P}^t_\beta(\varphi)}$$

where $\mathbb{P}^t_\beta(\varphi)$ is a probability distribution over the space of commitments that the next commitment β will make to α is φ. Similarly, for an overall estimate of β's *reputation* to α: $M(\alpha,\beta) = 1 - \sum_\varphi \mathbb{P}^t_\beta(\varphi)\,[1 - M(\alpha,\beta,\varphi)]$.

Preferred enactments. The previous measure, 'Ideal enactments', requires that an ideal distribution, $\mathbb{P}^t_I(\varphi'|\varphi, e)$, has to be specified for each φ. Here we measure the extent to which the enactment φ' is preferable to the commitment φ. Given a predicate

$\text{Prefer}(c_1, c_2, e)$ meaning that α prefers c_1 to c_2 in environment e. An evaluation of $\mathbb{P}^t(\text{Prefer}(c_1, c_2, e))$ may be defined using $\text{Sim}(\cdot)$ and the evaluation function $\boldsymbol{w}(\cdot)$ — but we do not detail it here. Then if $\varphi \leq o$:

$$M(\alpha, \beta, \varphi) = \sum_{\varphi'} \mathbb{P}^t(\text{Prefer}(\varphi', \varphi, o))\mathbb{P}^t(\varphi' \mid \varphi)$$
$$M(\alpha, \beta, o) = \frac{\sum_{\varphi:\varphi \leq o} \mathbb{P}^t_\beta(\varphi) M(\alpha, \beta, \varphi)}{\sum_{\varphi:\varphi \leq o} \mathbb{P}^t_\beta(\varphi)}$$

Certainty in enactment. Here we measure the consistency in expected acceptable enactment of commitments, or "the lack of expected uncertainty in those possible enactments that are preferred to the commitment as specified". If $\varphi \leq o$ let: $\Phi_+(\varphi, o, \kappa) = \{\varphi' \mid \mathbb{P}^t(\text{Prefer}(\varphi', \varphi, o)) > \kappa\}$ for some constant κ, and:

$$M(\alpha, \beta, \varphi) = 1 + \frac{1}{B^*} \cdot \sum_{\varphi' \in \Phi_+(\varphi, o, \kappa)} \mathbb{P}^t_+(\varphi' | \varphi) \log \mathbb{P}^t_+(\varphi' | \varphi)$$

where $\mathbb{P}^t_+(\varphi' | \varphi)$ is the normalisation of $\mathbb{P}^t(\varphi' | \varphi)$ for $\varphi' \in \Phi_+(\varphi, o, \kappa)$,

$$B^* = \begin{cases} 1 & \text{if } |\Phi_+(\varphi, o, \kappa)| = 1 \\ \log |\Phi_+(\varphi, o, \kappa)| & \text{otherwise} \end{cases}$$

5 Semantic Alignment

Information-based agents treat *everything* in the Semantic Web as uncertain — including the meaning of terms expressed in other agents' ontologies. They model their uncertain beliefs using random variables and probability distributions that are in a constant state of decay (as described in Section 3.1) and incorporate incoming information using entropy-based inference (as described in Section 3.3).

This discussion is from the point of view of agent α with ontology A who receives an illocution from agent β containing a term c expressed in β's ontology B. We assume that there is a term in A that corresponds precisely to $c \in B$.[6] Let $\Psi_{AB}(c)$ denote the term in ontology A that corresponds to $c \in B$, and let $F_{AB}(c)$ be a random variable over a subset of A representing α's beliefs about the meaning of c. $F_{AB}(c)$ may be defined at some level of abstraction determined by the $\leq$ relation — ie: not necessarily at the lowest level on the *is-a* hierarchy. In the absence of any information, the probability distribution for $F_{AB}(c)$ will be a maximum entropy distribution. We now consider how α can increase the certainty in $F_{AB}(c)$.

Suppose that α has signed a contract with β and that β's commitment b is described as having property $d \in B$. When β enacts that commitment, as b', it may not necessarily be as promised, and, as described in Section 2, ξ will advise α accurately of b'

[6] This assumption can be avoided by simply adding another layer to the analysis and constructing a probability distribution for c across the space $\{$may be represented in A, may not be represented in $A\}$ — we ignore this complication.

in terms of A, d'. So α will have the evidence that α promised $d \in B$ and delivered $d' \in A$. This evidence can be used to reduce the uncertainty of $F_{AB}(c)$ using one of the methods described in Section 3.4. Another means of reducing the entropy $\mathbb{H}(F_{AB}(c))$ is through dialogue with β, or any other agent, who may communicate information about the meaning of d. α will temper any such advice in line with the estimated trustworthiness of the agent as described in Section 4, and with an estimate of the agent's reliability in this context as described in Section 3.3, and will then permit the qualified advice to update $F_{AB}(c)$ as long as the condition in Equation 3 is satisfied. We have described how α reduces $\mathbb{H}(F_{AB}(c))$ following the receipt of messages — we now discuss proactive action that α may take to achieve this goal.

A simple example of proactive behaviour is to ask questions that require a "yes/no" response. This leads to the issues of which question to ask, and to whom should the question be directed? A simple, but powerful, strategy is to construct the question whose answer will yield maximum reduction in entropy — ie: maximum information gain. These estimates are made in reference to the $J_s^{X_i}$("yes/no") update functions that determine the affect that each illocution has on the set of distributions that make up the world model. The update functions $J_s^{F_{AB}(c)}(\cdot)$ have to do two separate jobs: (i) they translate a message expressed in B into an expectation in terms of α's ontology A, then (ii) using this expectation they induce constraints on α's world model derived from that expectation. The update functions are the only means of updating the world model, and so they enable us to identify information that has the capacity to reduce the entropy to an acceptable level.

6 Interaction Models

From an agent perspective, the co-ordination of services is naturally seen as a complex process management problem — an area where agent technology has proved itself. We do not consider mobile agents due to security concerns — in any case, the value of mobility in process management is questionable as the agents are typically large. A conventional multiagent system that tracks both the process constraints (generally time and cost) and the process ownership[7] is eminently suitable. Constraints are essential to managing processes that are co-ordinating the quality of service delivery — where sub-processes have, and maintain, a budget that is passed from agent to agent with the process. The processes involved in the co-ordination of services will be unpredictable and prone to failure. This indicates that they should be formalised as *goal-driven* processes [15], that is they should be conceptualised as agent plans expressed in terms of the goals that are to be achieved. This abstraction enables *what* is to be achieved and *when* that has to happen to be considered separately from *how* it will be achieved and *who* will do the work. Here each high-level goal has at least one plan whose *body* is a state-chart of goals, and atomic goals are associated with some procedural program. To cope with plan failure at each level in this framework, each plan has three exits: success (✓), fail (✗) and abort(**A**), with appropriate associated conditions and actions. This

[7] If an agent sends a process off into a distribute multiagent system then it relinquishes immediate control over it. To maintain this control the whole system abides by an ownership convention so that the agent who initiates a process will not "loose" control of it.

means that the plans are near-failure-proof but are expensive to construct. This expense is justified by the prospect of reusing plans to manage similar processes.

Dialogical interaction takes place not simply for the purpose of clarification but to reach some sort of service level agreement or contract — that could be a contract deliver a service with particular characteristics. Given two contracts δ and δ' expressed in terms of concepts $\{o_1, \ldots, o_i\}$ and $\{o'_1, \ldots, o'_j\}$ respectively, the (non-symmetric) distance of δ' from δ is given by the vector $\boldsymbol{\Gamma}(\delta, \delta') = (d_k : o''_k)_{k=1}^{i}$ where $d_k = \min_x\{\mathrm{Sim}(o_k, o'_x) \mid x = 1, \ldots, j\}$, $o''_k = \sup(\arg\min_x\{\mathrm{Sim}(o_k, x) \mid x = o'_1, \ldots, o'_j\}$ $, o_k)$ and the function $\sup(\cdot, \cdot)$ is the supremum of two concepts in the ontology. $\boldsymbol{\Gamma}(\delta, \delta')$ is a simple measure of how dissimilar δ' is to δ and enables α to metricate contract space and to "work around" or "move away from" a contract under consideration. Every time an agent communicates it gives away information. So even for purely self-interested agents, interaction is a semi-cooperative process. If agent β sends α a proposed contract δ then the information gain observed by α, $\mathbb{I}(\alpha, \beta, \delta)$, is the resulting reduction in the entropy of α's world model $\mathcal{M}$. An approach to issue-tradeoffs is described in [5]. That strategy attempts to make an acceptable offer by "walking round" the iso-curve of α's previous proposal towards β's current proposal δ. We extend that idea here, and respond to δ with a proposed contract δ' that should be optimally acceptable to β whilst giving β equitable information gain: $\arg\max_{\delta'}\{\ \mathbb{P}^t(\mathrm{acc}(\beta, \alpha, \delta')) \mid \mathbb{I}^{t-1}(\alpha, \beta, \delta) \approx \mathbb{I}^t(\beta, \alpha, \delta')\ \}$, where the predicate $\mathrm{acc}(\cdot)$ is as described in Section 3.2. This strategy aims to be *fair* in revealing α's private information. Unlike quasi-utilitarian measures, both the measure of information gain and the $\boldsymbol{\Gamma}(\cdot, \cdot)$ measure apply to *all* illocutions.

7 Conclusion

Information-based agency meets the demands identified for agents and the Semantic Web. The agents' communication language is quite general and accommodates on each agent's ontology (Section 2.1). These agents treat everything in their world as uncertain and model this uncertainty using probability distributions. They employ minimum relative entropy inference to update their probability distributions as new information becomes available (Section 3.2) — information-based agency manages uncertainty using probability theory and Bayesian inference. These agents deal with semantic alignment, and with *all* interaction, as an uncertainty reducing exercise until the cost of further reduction of uncertainty out-weighs expected benefits (Section 5). The interaction model for the co-ordination of services capitalises on the track-record of multiagent systems in process management applications, and models service co-ordination as a time-constrained, resource-constained process management problem — the co-ordination of services is then achieved by managing these processes as goal-driven processes (Section 6). The estimation of trust, that is so important to the Semantic Web, is achieved here with a computationally grounded method that is unified with the estimation of reliability of information and the reputation of information providers. The information-based agency project is on-going — recent work [16] describes how agents can build working relationships in the information world.

Acknowledgements. Carles Sierra's research is supported by the OpenKnowledge STREP project, sponsored by the European Commission under contract number FP6-027253.

References

1. Sierra, C., Debenham, J.: Information-based agency. In: Proceedings of Twentieth International Joint Conference on Artificial Intelligence IJCAI-07, Hyderabad, India (2007) 1513–1518
2. Berners-Lee, T., Hall, W., Hendler, J.A., OHara, K., Shadbolt, N., Weitzner, D.J.: A framework for web science. Foundations and trend in Web Science **1** (2006) 1–130
3. MacKay, D.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2003)
4. Jennings, N., Faratin, P., Lomuscio, A., Parsons, S., Sierra, C., Wooldridge, M.: Automated negotiation: Prospects, methods and challenges. International Journal of Group Decision and Negotiation **10** (2001) 199–215
5. Faratin, P., Sierra, C., Jennings, N.: Using similarity criteria to make issue trade-offs in automated negotiation. Journal of Artificial Intelligence **142** (2003) 205–237
6. Rosenschein, J.S., Zlotkin, G.: Rules of Encounter. The MIT Press, Cambridge, USA (1994)
7. Kraus, S.: Negotiation and cooperation in multi-agent environments. Artificial Intelligence **94** (1997) 79–97
8. Arcos, J.L., Esteva, M., Noriega, P., Rodríguez, J.A., Sierra, C.: Environment engineering for multiagent systems. Journal on Engineering Applications of Artificial Intelligence **18** (2005)
9. Sierra, C., Debenham, J.: Trust and honour in information-based agency. In Stone, P., Weiss, G., eds.: Proceedings Fifth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2006, Hakodate, Japan, ACM Press, New York (2006) 1225 – 1232
10. Kalfoglou, Y., Schorlemmer, M.: IF-Map: An ontology-mapping method based on information-flow theory. In Spaccapietra, S., March, S., Aberer, K., eds.: Journal on Data Semantics I. Volume 2800 of Lecture Notes in Computer Science. Springer-Verlag: Heidelberg, Germany (2003) 98–127
11. Li, Y., Bandar, Z.A., McLean, D.: An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on Knowledge and Data Engineering **15** (2003) 871 – 882
12. Cheeseman, P., Stutz, J.: On The Relationship between Bayesian and Maximum Entropy Inference. In: Bayesian Inference and Maximum Entropy Methods in Science and Engineering. American Institute of Physics, Melville, NY, USA (2004) 445 – 461
13. Paris, J.: Common sense and maximum entropy. Synthese **117** (1999) 75 – 93
14. Debenham, J.: Bargaining with information. In Jennings, N., Sierra, C., Sonenberg, L., Tambe, M., eds.: Proceedings Third International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2004, ACM Press, New York (2004) 664 – 671
15. Debenham, J.: A multiagent system manages collaboration in emergent processes. In Dignum, F., Dignum, V., Koenig, S., Kraus, S., Singh, M., Wooldridge, M., eds.: Proceedings Fourth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2005, Utrecht, The Netherlands, ACM Press, New York (2005) 175 – 182
16. Sierra, C., Debenham, J.: The LOGIC Negotiation Model. In: Proceedings Sixth International Conference on Autonomous Agents and Multi Agent Systems AAMAS-2007, Honolulu, Hawai'i (2007)