Tracking Reliability and Helpfulness in Agent Interactions

Min Xu, Lin Padgham*, Aloys Mbala, James Harland
School of Computer Science and Information Technology
RMIT University
GPO Box 2476V
Melbourne, 3001
Australia
Phone: +613 9925 2348
Fax: +613 9662 1617
Email: {*min,linpa,aloys,jah*}*@cs.rmit.edu.au*

February 23, 2006

Abstract

A critical aspect of open systems such as the Internet is the interactions amongst the component agents of the system. Often this interaction is organised around social principles, in that one agent may request the help of another, and in turn may make a commitment to assist another when requested. In this paper we investigate two measures of the social responsibility of an agent known as *reliability* and *helpfulness*. Intuitively, reliability measures how good an agent is at keeping its commitments, and helpfulness measures how willing an agent is to make a commitment, when requested for help. We discuss these notions in the context of FIPA protocols. It is important to note that these measures are dependent only on the messages exchanged between the agents, and do not make any assumptions about the internal organisation of the agents. This means that these measures are both applicable to any variety of software agent, and externally verifiable, i.e. able to be calculated by anyone with access to the messages exchanged.

Keywords: Social commitment, reliability, helpfulness, agent protocols, measurement.

*All correspondence should be addressed to this author.

1 Introduction

Open environments such as the Internet pose new challenges for software systems because of their dynamic nature and the heterogeneity of entities in them. Agent-based systems are becoming increasingly important for the development of applications in such environments because of their ability to cope with uncertainty and to adapt their behaviour to constantly changing conditions.

A critical aspect of the development of such systems is the manner in which agents cooperate. In order to organise interactions between autonomous agents in open systems, it is common to utilise concepts such as *trust* [RHJ04] and *commitment* [YS01]. There have been a number of papers written on various aspects of commitments [Sin91, YS01, FC02, FC04]. The particular aspect of interest in this paper is *social commitment*, i.e. an undertaking given by one agent to another (in contrast with the *single-minded* or *open-minded* commitment of [RG91], which is concerned with how a single agent may come to change its goals). However, an agent which makes a commitment to another must clearly reflect this in its own goals and internal reasoning processes. In this paper we are mainly concerned with the behaviour of an agent in dealing with its social commitments. Conceptually, a social commitment is an undertaking given by one agent (the *debtor*) to another (the *creditor*) for which the main incentive to fulfil the commitment is social pressure from the society of agents. In order to exert such pressure, it is clearly necessary for the agent society as a whole to be able to make objective judgements about the commitments of an agent and whether they have been kept or not. Accordingly, such commitments must be observable by all agents; we ensure this by requiring that commitments be established by an exchange of messages.[1]

With most agent systems, the behaviour of the environment is impossible to predict, and so the only guarantee that an agent can give to another party is about its own actions; anything else is beyond its control, particularly the actions of other agents. In addition, agents are assumed to be self-interested and to work to maximise their own utility. Although such environments, by their very nature, may appear to mitigate against any form of cooperative interaction, Jennings and Campos[JC97] have shown that socially responsible agents can lead to both good system-wide performance and good individual performance when they are placed in a social context. Furthermore, a socially rational agent may take into account the effect of its actions on the community as a whole, rather than just the effect on its own goals [HJ97]. The most common form of interaction between agents is requesting help from other agents and responding to such requests [ACR96].

The two concepts which we believe are critical to the success of such interactions are *reliability* and *helpfulness*. Reliability is a measure of the extent to which an agent keeps its commitments, while helpfulness is a measure of the extent to which an agent is willing to make commitments. When making decisions about interactions in an open environment, an agent can use these characteristics to evaluate potential interaction partners (similar to the credibility ratings used in on-line auction systems such as eBay). In our framework, these are based purely on the agent's previous history of interactions, and so the reliability and helpfulness of an agent (i.e. the agent's *social responsibility*) can be measured by observing the external interactions of agents with each other. This provides a simple and objective means of evaluation for systems used for elecronic trading, contract negotiation or composition of web services. As measurements are made purely on information available to all agents, there is no scope for agents to make unsubstantiated claims for their reliability. Note that we do not consider issues of identity or intention; we assume that all agents in the system are identified, and act honestly. Establish-

[1]For the remainder of this paper, we will use *commitment* to mean *social commitment* unless indicated otherwise.

ing an agent's identity and detecting subversive behaviour are topics of interest, but are outside the scope of this paper.

Ideally, from the society viewpoint, an agent would be both maximally reliable and maximally helpful. However, concern about being very reliable may make an agent overly cautious in responding to requests; an agent that never makes a commitment can never fail to fulfil its commitments. On the other hand, a very helpful agent may take on too many commitments, and then find that it is unable to fulfill them all, thus leading to low reliability. Both extreme cases — agents that always make commitments, but seldom manage to keep them, or agents that seldom or never make commitments — are useless in practice. Hence there will always be a trade-off between reliability and helpfulness.

An important aspect of both of these measures are that they can be evaluated purely by external means. We only need to know what requests are made to a given agent, and what commitments it undertakes, both of which are given by messages exchanged between agents. This means that our measures of reliability and helpfulness do not depend on the internal architectures of the agents. We do assume, though, that there are some specified interaction protocols, which allow an observer to determine when commitments are requested, made, fulfilled or broken. We also assume that the agents follow the specified protocols. In order to provide a point of technical focus for examples of our techniques, we concentrate on the interaction protocols specified by FIPA [fIPAFb], the agents standards body. However, our methods are not limited to these protocols, provided that it is clear how to identify the contents of a message (and in particular a message indicating acceptance of a commitment).

This property of external verification is the main point of difference between our work and trust models [ARS+05, SS02]. Being based purely on the visible history of interaction, there is no scope for third-party recommendations, nor for any inference based on social networks. In other words, what we provide as an input to each agent's internal reasoning process is a simple, concrete and verifiable measure of reliability based on past history. Related notions such as reputation provide more sophisticated mechanisms, but we believe that our approach provides a simple platform on which sublter forms of inference can be based.

The rest of the paper is organized as follows. In Section 2 we discuss the concepts of reliability and helpfulness, and in Section 3 we discuss how to measure these, including a combined measure which we call *responsibility*. Section 4 discusses how these measurements can be applied to the specific FIPA interaction protocols, and provides and algorithm suitable for use in a FIPA compliant platform. Section 5 discusses the trade-offs between reliability and helpfulness, and how an agent might reason about these. Section 6 discusses related work and Section 7 summarizes our results and conclusions as well as indicating possibilities for further work.

2 Defining Reliability and Helpfulness

Our conceptual point of departure is that of a society of agents. Each agent is assumed to be self-interested, and able to pursue its own goals. However, an agent may request assistance from another agent by sending a *request* message. This request can be either ignored, denied or accepted by the other agent, and in the latter case we say the agent has made a *commitment* to the first agent to perform the requested task (and thus the commitment has been *established*). In this sense all commitments arise from messages exchanged between agents.

For example, in the Contract Net protocol [Smi80], the initiator agent will call for bids to perform a certain task, select one or more winning bidders, and assign the task to the relevant agent(s). Once the task is assigned to an agent making a bid, the bidding agent acquires a commitment to perform the allotted task.

In this paper we assume that once a commitment is made, one of three things happen:

- The debtor agent successfully completes the allocated task, and sends a message to the creditor, indicating it has been *successful*[2]. In this case we say the commitment has been *fulfilled*.
- The debtor agent fails to complete the allocated task, and sends a *failure* message to the creditor. In this case we say the commitment has been *broken*.
- The creditor agent cancels the commitment by sending a *cancel* message to the debtor. In this case we say the commitment has been *cancelled*.

Note that the third case above is the only way in which a commitment may be cancelled; in particular, the debtor has no way to cancel a commitment. In practice, there is a fourth case as well, in which the outcome is not yet known; we call such commitments *pending*. This point is discussed further in the next section.

As we have discussed, a reliable agent is one which keeps the commitments it has made. Ideally, we may wish to define an agent as reliable only when it fulfills every commitment that it makes. However, in practice we cannot always expect this, as even the most reliable agent may sometimes fail. Moreover, we want to be able to compare the relative reliabilities of agents which fail to keep their commitments.

Accordingly, we define the reliability of an agent as the ratio of its fulfilled commitments to its established commitments. Note that we consider this value undefined if there are no such commitments.

Helpfulness is defined similarly, i.e. as the ratio of the commitments established to the requests received. This value is also undefined if the agent does not receive any requests.

In some cases, it may be useful to measure not just the overall reliability and helpfulness of an agent, but particular subsets of the overall interactions of the agent. For example, in the Contract Net protocol, the manager agent may be less interested in the overall reliability of a particular contractor than its reliability on tasks requested by the manager (which may reflect a deliberate policy on the contractor's part to prioritise tasks requested by the manager over other requests). For this reason we distinguish between *global reliability*, (i.e. reliability measured on all known interactions) and *relative reliability* (i.e. reliability measured on interactions only with a particular group of creditors). Similar remarks apply to helpfulness.

3 Measuring Reliability and Helpfulness

3.1 Measuring Reliability

In the previous section we have established the general notion of reliability as the extent to which an agent fulfills its commitments, while helpfulness is the extent to which an agent is willing to make commitments. However, if we are to actually measure these characteristics we must clarify a number of details. While the general principle is that we want the ratio of fulfilled commitments to established commitments, we must also determine how to deal with such things as cancelled commitments and pending commitments. If an agent is released from a commitment by a cancellation, then this should not have any negative impact on the reliability of that agent. However, it also shouldn't be considered as a fulfilled commitment. This leads us to the notion that it is not the ratio of fulfilled commitments to established commitments

[2] This message may contain results of the activity committed to, or may simply indicate successful completion.

which is of interest, but rather the ratio of fulfilled commitments to what we will call *valid commitments*, that is commitments which are made, and not cancelled.

At any point in time when we wish to assess the reliability of an agent, we will in essence be looking at that agent's reliability over some particular time period. It is likely that within that time period some commitments will be pending - they will have been established, but have neither been fulfilled nor broken. Using the ratio of fulfilled commitments to established commitments assumes that all pending commitments will be broken; using the ratio of fulfilled commitments plus pending commitments to established commitments assumes that all pending commitments will be fulfilled. Neither is likely to be a fair assumption. It seems more reasonable to "distribute" the pending commitments according to the current ratio of fulfilled commitments to established commitments, thus using the previous history as a guide. Consequently we define a notion of *finalised commitments* which are valid commitments which have been completed, i.e. they have been either fulfilled or broken. Our reliability measure then can be given as:

$$Reliability(p) = \begin{cases} \dfrac{\|FulfilledComm(p)\|}{\|FinalisedComm(p)\|} & if \|FinalisedComm(p)\| > 0 \\ \\ Undefined & \text{Otherwise} \end{cases}$$

This definition is equivalent to the more complex formula, which assumes that pending commitments are fulfilled or broken in the same proportion as the currently finalised commitments. As finalised commitments are simply the sum of fulfilled and broken commitments, this definition also has the advantage that it is not necessary to measure or track established or cancelled commitments (although of course we can always determine the number of finalised commmitments from the number of established and pending ones, if desired).

For example, consider an agent which has previously established 100 commitments, of which 82 were fulfilled, 13 broken and 5 cancelled, and which has 10 pending commitments. Thus we have 110 established commitments (the 100 previously completed and the 10 pending), and 95 finalised commitments. Hence its reliability is 82/(82+13) = 86%. The more complex formula mentioned above would give (82 + (82/95)*10)/(82+13+10) = (82 + 8.6)/105 = 86%. Note that the first calculation does not require the number of pending commitments to be known; only the fulfilled and broken ones, i.e. only those whose status will not change (unlike pending commitments, whose final status may be any of fulfilled, broken or cancelled).

3.2 Measuring Helpfulness

As discussed above, helpfulness is a measure of the willingness of an agent to establish commitments. As with reliability, there may also be pending situations (i.e. where an agent has not yet replied to a request), and again it seems appropriate to distribute the pending outcomes according to the agent's previous history of responses. Consequently we define our helpfulness measure using *positive responses* to requests, as a proportion of *total responses* to requests, i.e. positive plus negative responses. The helpfulness measure is thus defined as:

$$Helpfulness(p) = \begin{cases} \dfrac{\|PositiveResponses(p)\|}{\|TotalResponses(p)\|} & if \|TotalResponses(p)\| > 0 \\ \\ Undefined & \text{Otherwise} \end{cases}$$

Alternatively we may wish to consider that in the situation of requests, and responses to requests, there is likely to be little time delay, and therefore we are not concerned with any distortion due to pending requests. In addition, it may often be the case that a failure to respond is in fact a negative response, or at least a failure to be helpful. In particular, an agent that never sends a negative response will have a helpfulness rating of 100% even if it does not take on many commitments. Consequently an alternative definition is:

$$Helpfulness(p) = \begin{cases} \dfrac{\|PositiveResponses(p)\|}{\|TotalValidRequests(p)\|} & if \|TotalValidRequests(p)\| ¿0 \\ Undefined & \text{Otherwise} \end{cases}$$

The main difference here is the information required to evaluate each one. The first requires only the outgoing messages of the given agent, (which, as above, the agent can manipulate if it so chooses) whereas the second needs to also look at the incoming messages of the agent, and hence is less prone to abuse at the cost of ignoring pending requests. Accordingly agents which either do not give negative responses or tend to reply more slowly to requests than others will tend to have lower helpfulness ratings, which seems appropriate.

It should be stressed that this particular measure of helpfulness is designed to be simple; there are a variety of others, such as incorporating the average time to respond, which could be used. More sophisticated measures will clearly provide a more informative picture, but at a corresponding cost. As our main focus is on providing simple and verifiable measures, we do not pursue more sophisticated measures in this paper.

With both of these measures, where there is no history, (i.e. the number of responses or finalised commitments is zero), then the reliability or helpfulness measure is undefined. In an implementation other decisions could be made when no specific information is available. For example, a default value (either fixed, or dynamic, such as the current average for agents in the system) could be assigned to agents without a history (such as a new agent being added).

Both these measures can readily be modified in the obvious way to provide a measure of reliability/helpfullness with respect to a particular agent or group of agents.

3.3 Combining Reliability and Helpfulness

Measurements of helpfulness and reliability allow us to compare agents on one or other of these characteristics. However, as has been noted previously, there is a tension between these two characteristics, as, in a demanding environment, greater helpfulness is likely to lead to lower reliability (when more tasks are taken on, there is a greater risk of conflict, or inadequate resources). We would like to have a single score, incorporating both helpfulness and reliability, on which we can compare agents.

We note that in many ways reliability and helpfulness exhibit the same kind of tension as that between *precision* and *recall*, the basic measures for evaluating search strategies in information retrieval [VR79]. In general, it is not possible to have perfect scores in both measures; better performance on one measure tends to be associated with worse performance on the other.

Figure 1 illustrates this inverse relationship.

The inverse relationship between recall and precision has been well understood for several decades [CMK66, RBJ89]. The notion of a harmonic mean, known as the F measure, has been used to combine precision and recall to produce a single score on which systems can be compared.

We define a similar F measure in our context as:

$$F = \frac{2 * \mathit{Reliability} * Helpfulness}{\mathit{Reliability} + Helpfulness}$$

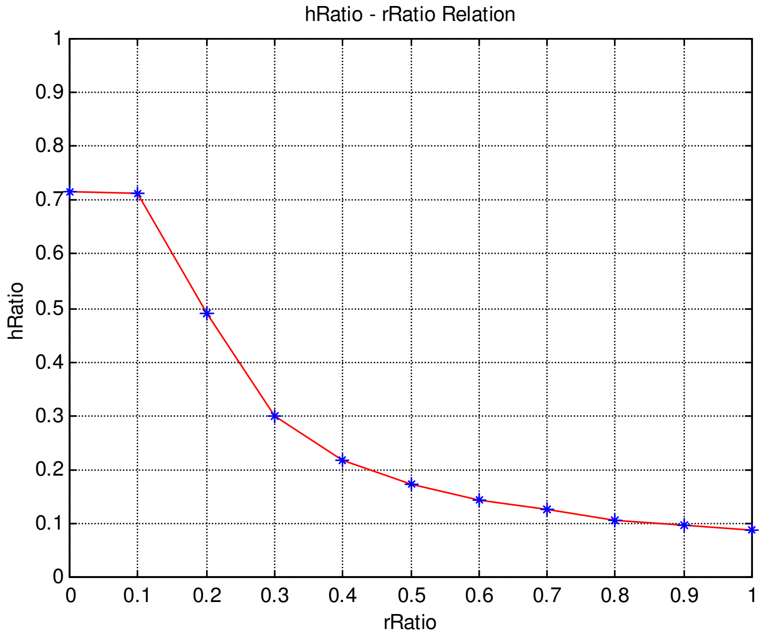

Figure 1: Helpfulness (hRatio) and reliability (hRatio) are inversely related.

We will call this combined measure *responsibility*, as it provides a measurement of the balance between helpfulness and reliability.

An important property of this measure is that it requires a balance between the two competing characteristics. Unlike a standard mean, it is not possible to compensate for a low value on one aspect with a high value on another. For example, comparing scores of 90 and 10, with 50 and 50, each would have a mean of 50, but the former has an F value of 18 while the latter has an F value of 50. Even a balanced score of 20 and 20 gives an improved F value of 20, as compared to the imbalance of 90 and 10.

Figure 2 shows how the F measure varies as its component values vary.

A variant of the F measure, again used in Information Retrieval, is the "E measure" which allows for differing weights to be placed on the two competing characteristics. We will call this *weighted responsibility* and it is given as:

$$E = \frac{(1 + \beta^2) * Reliability * Helpfulness}{\beta^2 * Reliability + Helpfulness}$$

where

- $\beta = 1$ gives equal weight to reliability and helpfulness (E=F)
- $\beta > 1$ gives greater weight to reliability
- $\beta < 1$ gives greater weight to helpfulness

Reliability and helpfulness are clearly distinct concepts which it can be useful to reason about individually. For instance in some situations it may be preferable to be assured of high reliability, and obtaining a sufficiently helpful agent to interact with is not an issue. In other situations, where demand is high and it is critical to establish a commitment quickly, helpfulness

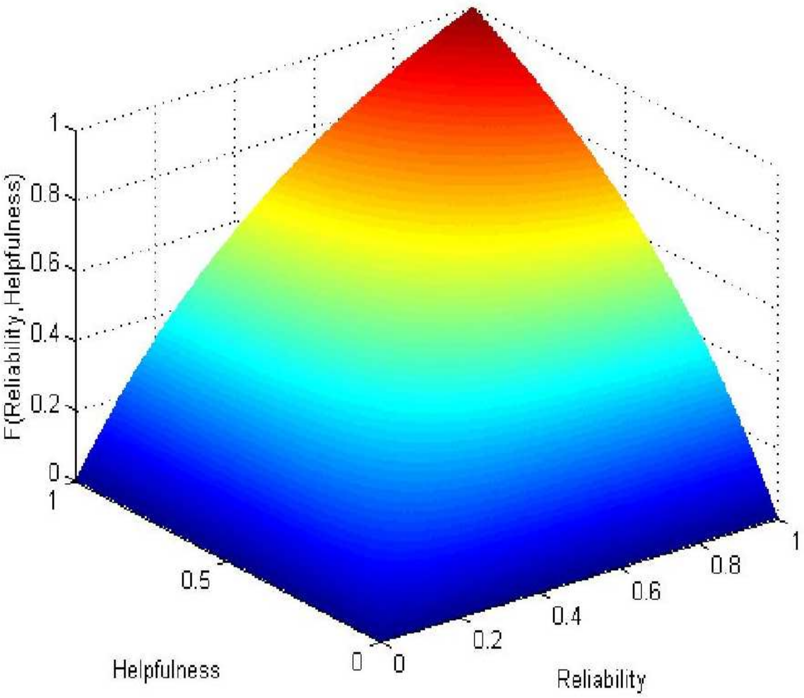

Figure 2: Figure showing the value of F as its component values change

may well be a more critical issue. However it is certainly useful to have a measure which allows us to compare agents across both these characteristics, in a simple but principled manner. The E and F measures provide this ability.

4 Using FIPA to Measure Helpfulness and Reliability

As discussed above, a key feature of our approach is that the measurements of reliability and helpfulness are determined by interactions between agents. In order for interactions to be the basis for measurement, it is necessary for there to be some level at which interactions can be noted and categorised without being concerned about the specific details of the message being sent. The FIPA (Foundation for Intelligent Physical Agents)[3] Communicative Act Library [fIPAFa] defines a set of message types (performatives) with specific semantics which allow us to categorise the type of an interaction in this way, without being concerned about the details of the actual message.

Some performatives implicitly assume certain social commitments of sending/receiving agents. Fornara and Colombetti have proposed a method for agent communication, where the meanings of messages are expressed by the social notion of commitments [FC02]. They analyze the evolution of social commitments and also give a formal semantics for Agent Communication Language (ACL) messages based on social commitments. However their work differs from ours in that their focus is on giving a commitment based semantics to ACLs, whereas our focus is on using the concept of commitment to measure characteristics of an agent interacting within

[3]FIPA is the standards body for agent systems.

a multi-agent system.

We have chosen to stay within the FIPA defined ACL, as it is widely used, and provides a mechanism whereby our suggestions can be put to immediate use. We assume that all agents properly follow the prescribed protocols. This assumption will of course be false whenever an agent leaves the space (perhaps by having its server go down), without properly terminating all protocols it is involved in. However our approach to essentially ignoring any pending interactions (at least in the case of responsibility) mitigates against these effects.

Within FIPA protocols, an agent can always respond with a *not-understood*, informing the other party that it did not understand what was communicated. This performative may terminate the interaction and possibly violate any commitments established. An agent initiating a request can also send a *cancel* at any stage, thus terminating any commitment established. These features complicate the analysis and measurement of reliability and helpfulness in real world situations. Note also that agents can be engaged in multiple protocols simultaneously.

The effect of the various kinds of exceptions that can occur, such as *not-understood, cancel* and *time-out*, vary depending on where in the protocol they arise. Consequently it is necessary, when measuring the helpfulness and reliability characteristics of specific agents, to track the individual conversations that occur. As a result, it is no disadvantage to specify the reliability and helpfulness semantics of particular performatives, within the context of specific protocols.

We have therefore analyzed the existing FIPA protocols to establish which performatives, in which protocols, are relevant for identifying the interactions appropriate to measuring helpfulness and reliability. We examine each of the nine accepted standard FIPA protocols, (Brokering, Contract Net, Iterated Contract Net, Propose, Query, Recruiting, Request, Request-when and Subscribe). Various decisions could be made about how to interpret the different performatives within each protocol. We examine a particular set of decisions to illuminate how the concepts can be interpreted. As a result of this analysis we also make some recommendations regarding well-structured protocols, with respect to measuring reliability and helpfulness. In Section 4.3 we also provide an algorithm for measuring reliability and helpfulness in some of the protocols.

4.1 Helpfulness Analysis within FIPA Protocols

The formula for Helpfulness was given previously in two alternative forms:

$$Helpfulness(p) = \begin{cases} \dfrac{\|PositiveResponses(p)\|}{\|TotalResponses(p)\|} & \text{if } \|TotalResponses(p)\| > 0 \\ Undefined & \text{Otherwise} \end{cases}$$

or

$$Helpfulness(p) = \begin{cases} \dfrac{\|PositiveResponses(p)\|}{\|TotalValidRequests(p)\|} & \text{if } \|TotalValidRequests(p)\| > 0 \\ Undefined & \text{Otherwise} \end{cases}$$

We must now consider how to observe PositiveResponses and either TotalResponses or TotalValidRequests, within each of the FIPA protocols, depending on which version of the helpfulness formula we choose to use.

We first examine how to interpret the notion of *request* within the various protocols. For the *Brokering, Query, Recruiting, Request, Request When* and *Subscribe* protocols, this is quite straightforward. The requests are simply the performatives *proxy, query-if, query-ref, request, request-when* , and *subscribe.* However for the remaining protocols it is not so straightforward.

For the *Propose* protocol the initiator is proposing to do something for the *participant* without any request having happened. One possibility could be to count the *propose* as a positive response, and also increment the number of requests, in order to maintain the balance. The other option is to not cover the *Propose* protocol with respect to helpfulness.

For the Contract Net Protocols[4], we choose to treat the *cfp* (call for proposal) as a request. We could alternatively have taken the view that these protocols are not ones in which an agent can demonstrate helpfulness.

However, as mentioned previously, we need to exclude in our measure of TotalValidRequests any requests which are cancelled before they have been responded to, or requests which are ill-formed and therefore not understood by the receiver.

Determining which performatives can be considered as positive responses seems relatively straightforward once we have established what a request is. For example *agree* in the *Brokering, Query, Request, Request When* and *Subscribe* protocols and *proposal* in the *Contract Net* protocols[5] are clearly Positive Responses. However the *Query, Request, Request When* and *Subscribe* protocols do not necessarily require an *agree* message - the *inform* may just be sent directly. Consequently when monitoring these protocols we must interpret either *agree* or *inform* (but not both in a single conversation) as indicators of Positive Response. Alternatively we can monitor Negative Responses and assume that Positive Responses are equal to Total Valid Requests minus Negative Responses, for those protocols where the *agree* is optional.

4.2 Reliability Analysis within FIPA Protocols

The formula for measuring reliability was given as:

$$Reliability(p) = \begin{cases} \dfrac{\|FulfilledComm(p)\|}{\|FinalisedComm\|} & \text{if } \|FinalisedComm(p)\| > 0 \\ Undefined & \text{Otherwise} \end{cases}$$

For the *Contract Net, Query* and *Request* protocols commitments are normally finalised with either an *inform* or a *failure*, representing fulfilled or broken commitments respectively. Finalised commitments are thus simply the sum of the number of *inform* and *failure* messages.

The measurement expression requires that it is possible to determine when commitments have been finalised. On analyzing the *Brokering* and *Recruiting* protocols we identify some potential difficulties in establishing when commitment has been fulfilled. In both these protocols the final steps of the protocol come not from the *Broker* or *Recruiter* which established the commitment, but rather from one or more of the *Participant* agents engaged by the *Broker* or *Recruiter*. One viewpoint on reliability would be to count the *inform* from the *Recruiter* or *Broker* indicating that appropriate *Participant* agent(s) had been engaged as being the appropriate fulfillment of the commitment. Alternatively one may wish to require the *Broker* or *Recruiter* to engage *Participant* agents which are themselves reliable, and therefore not consider the commitment as fulfilled until one or more *Participant* agents send an *inform* or *reply-to* with the appropriate conversation id. In the *Recruitment* protocol this is further complicated by the fact that both the *inform* and the *reply-to* may go to a designated third party. Nevertheless these messages could be monitored using the conversation id. The simpler approach is simply to consider the commitment fulfilled when suitable *Participant* agents have been engaged.

[4]This can also be applied to the Dutch and English Auction protocols, which have however not yet been accepted as standard.

[5]Also applicable to the auction protocols.

In the *Subscribe* protocol, the commitment is to notify the subscriber whenever a relevant change occurs. In this case broken commitments would be cases where relevant change occurred but no notification was sent. Without access to information about the referenced object it is impossible to know whether or not the agent is reliable in informing all necessary changes. It is not possible for either a generic system monitor, or the agent who is the recipient of the commitment, to ascertain whether there may be commitment failures. However, a simplifying assumption is that when the first subscription is recieved (via an *inform* we can consider the commitment to be fulfilled. Consequently we count the first *inform* in a subscription as evidence of reliability.

A related issue is when the commitment made is actually a *conditional commitment* [YS01] This case occurs in the *Request When* protocol, which requires an action when a certain condition is true. Without access to both the content of the condition, and the world of the agent in which it is to be evaluated, it is impossible to ascertain when the action committed to should be performed, and thus when there is failed commitment.

This problem arises in any protocol where there is a condition attached, leading to a conditional commitment. If such a commitment is not fulfilled, it is impossible to determine if it is because of unreliability of the agent or because the preconditions of the commitment were not satisfied. Consequently, in this work we assume that there are no conditions buried in the content of messages. The *Request When* protocol is consequently excluded from consideration with respect to reliability.

The *Propose* protocol has already been discussed as somewhat unusual in that there is no request for assistance. Nevertheless, it can be assumed that when the *Initiator* sends the *propose* it is making a commitment to do the proposed action if the *Participant* agrees. However, despite the fact that in the description of the protocol it says "Completion of this IP with an *accept-proposal* act would typically be followed by ... the return of a status response", no such response is required within the protocol. Without this it is impossible to ascertain whether the *Initiator* who has established the commitment, has behaved reliably. Consequently we also exclude the *Propose* protocol from consideration of reliability. It is suggested that in reviewing protocols for standardisation it would be useful to consider the issue of observability of commitments undertaken and their fulfillment. Addition of an *inform* or *failure* in the *Propose* protocol, similar to other protocols would have enabled reliability to be observed.

We summarise our suggestions as to which messages to count in assessing helpfulness and reliability in Table 1.

FIPA Protocol	Performatives for Helpfulness		Performatives for Reliability	
	Request	Positive Response	Failed commitment[a]	Fulfilled commitment
Brokering[b]	proxy	agree	failure	inform
Contract Net Iterated Contract Net	cfp	propose	failure	inform
Propose	N/A[c]	propose	N/A[d]	N/A
Query	query-if query-ref	agree inform	failure	inform
Recruiting[e]	proxy	agree	failure	inform
Request	request	agree inform	failure	inform
Request-when	request	agree inform	N/A[f]	N/A
Subscribe	request	agree inform	failure	inform

Table 1: Summarizing helpfulness and reliability in FIPA standard protocols

[a]Finalized commitments are the sum of Failed and Fulfilled Commitments.
[b]We assume the option where the commitment is discharged once the proxy has been found.
[c]Assume an implicit request when propose observed.
[d]Reliability cannot be measured in this protocol as no message is specified.
[e]We assume the option where the commitment is discharged once the proxy has been found.
[f]Reliability cannot be measured due to conditional commitment, and inability to observe the condition.

4.3 Algorithm

We describe here an algorithm which will enable a process with access to the interactions in the system to calculate (global) reliability and helpfulness. Such a process could reasonably be attached to the Agent Management System (AMS) of a FIPA platform.

An alternative would be to have all agents compute values for the agents that they interact with, and then combine these values periodically, or whenever a global measure was required. However this would require that all agents to have such capabilities, and also introduces trustworthiness issues and the extra complexities of measuring relative scores. For simplicity we choose an algorithm which can be located in a central monitoring process. The measurement of relative reliability and helpfulness would naturally be distributed.

In order to measure helpfulness we will use the number of Valid Requests of agent i (VR_i), and the number of Positive Responses (PR_i). Similarly we count the number of Established Commitments (EC_i) and Fulfilled Commitments (FC_i) for measuring reliability.

In the FIPA protocols, a request is established whenever an agent initiates an interaction protocol by sending one of the list of performatives including *request, query-if, query-ref, request-when, cfp,* etc. This request will be valid unless the initiator sends a *cancel* performative immediately or the receiver agent replies with *not-understood.* Thus the validity of a request and the establishment of a commitment depend on the progression of the interaction protocol. Hence we define a double-index array VR_{ij}, in which VR_{ij} is a boolean value indicating whether a request received by agent i in the interaction protocol j is valid or not. Similar remarks apply to PR_{ij}, EC_{ij}, FC_{ij} regarding accepted requests, established commitments and fulfilled commitments respectively.

Note that there are three stages relevant to the calculation of helpfulness:

1. Either no request has been received or a request has been received and cancelled
2. A request has been received, but no response has been sent
3. A request has been received and a response (either positive or negative) has been sent

Once the third stage has been reached, it is then meaningful to calculate helpfulness. Similarly, there are three stages relevant to the calculation of reliability:

1. Either no request has been received or a request has been received and cancelled
2. A request has been received, but no response has been sent
3. A request has been received and a positive response has been sent (and so a commitment has been made).

Again, once the third stage has been reached, it is then meaningful to calculate reliability. Hence we keep track of each of these values with the variables $Consider\_Help_j$ and $Consider\_Rel_j$, each of which is initially zero. Consider the FIPA Request Interaction Protocol j. When an initiator sends a *request* performative, this will lead potentially to a valid request and also to an established commitment, and so we set the two variables to 1. If the participant replies with *agree* or *refuse*, we increment $Consider\_Help_j$. If the response is *agree*, then we also increment $Consider\_Rel_j$. If the response is *refuse*, then we decrement $Consider\_Rel_j$ (so that $Consider\_Rel_j$ is now 0). If the initiator sends a *cancel* performative, the we set $Consider\_Help_j$ to 0. Note that it is also possible for an agent to accept the request without sending *agree* to the initiator; we accordingly set $Consider\_Help_j$ and $Consider\_Rel_j$ when an *inform* or *failure* occurs.

We consider VR_{ij} and EC_{ij} in the computation of VR_i and EC_i only when $Consider\_Help_j$ and $Consider\_Rel_j$ are greater than 1. We denote this by

$$\sum_{Consider\_Help_j>1} VR_{ij} \quad \text{and} \quad \sum_{Consider\_Rel_j>1} EC_{ij}$$

Figure 3 shows the algorithm to measure the helpfulness and reliability of all agents known in the system for a given time interval. If a message has its receiver or its sender unknown in the system, then the list of known agents is updated. The algorithm is based on the assumption that the protocol in which each message was sent is known; otherwise, the message is simply discarded.

Figure 4 shows how to set the variables in the particular case of FIPA Request Interaction Protocol (this is used in place of the line "Set values for variables" above). Due to space limitations, we do not depict this for all FIPA Interaction Protocols. The algorithm considers the sequence of the interaction protocol and also the exceptions caused by *cancel* and *not-understood* performatives.

Data Structure:

i	a counter, $1 \leq i \leq N$
N	the number of agents in the system
$Agent_i$	the name of agent of rank i
VR_i	the number of valid requests received by $Agent_i$
AR_i	the number of accepted requests for $Agent_i$
EC_i	the number of commitments established with $Agent_i$
FC_i	the number of commitments fulfilled by $Agent_i$
j	contains the Conversation ID of the protocol
VR_{ij}	is a request received by $Agent_i$ in protocol j valid or not
AR_{ij}	is a request received by $Agent_i$ in protocol j accepted or not
EC_{ij}	is a request received by $Agent_i$ in protocol j an established commitment or not
FC_{ij}	is a commitment made by $Agent_i$ in protocol j fulfilled or not
$Consider\_Help_j$	the stage of protocol j at which a valid request should be counted
$Consider\_Rel_j$	the stage of protocol j at which an established commitment should be counted
H_i and R_i	helpfulness and reliability of $Agent_i$

Algorithm:

For each agent i
    For each message
        Extract protocol, j=conversation_id, sender, receiver;
        j $\leftarrow$ conversation_id;
        If (protocol is null) **Discard** the message;
        If (receiver or sender unknown in list of agents) **Update** $Agent_i$, N;
        Set $VR_{ij}, AR_{ij}, EC_{ij}, FC_{ij}, Consider\_Help_j$, $Consider\_Rel_j$ as appropriate for the protocol;
    EndFor
    $VR_i \leftarrow \Sigma_{Consider\_Help_j > 1} VR_{ij}$; $EC_i \leftarrow \Sigma_{Consider\_Rel_j > 1} EC_{ij}$;
    $AR_i \leftarrow \Sigma_j AR_{ij}$; $FC_i \leftarrow \Sigma_j FC_{ij}$;
    $H_i \leftarrow AR_i / VR_i$ if $VR_i > 0$;
    $R_i \leftarrow FC_i / EC_i$ if $EC_i > 0$;
EndFor

Figure 3: Measuring helpfulness and reliability.

Setting Variables:

If (PERFORMATIVE = request AND RECEIVER = $Agent_i$)
$VR_{ij} \leftarrow 1$; $EC_{ij} \leftarrow 1$; $Consider\_Help_j \leftarrow 1$; $Consider\_Rel_j \leftarrow 1$;
If (PERFORMATIVE = agree AND SENDER = $Agent_i$)
$Consider\_Help_j \leftarrow 1$; $Consider\_Rel_j \leftarrow Consider\_Rel_j + 1$;
If (PERFORMATIVE = refuse AND SENDER = $Agent_i$)
$AR_{ij} \leftarrow 0$; $EC_{ij} \leftarrow 0$; $Consider\_Help_j \leftarrow Consider\_Help_j + 1$; $Consider\_Rel_j \leftarrow 0$;
If (PERFORMATIVE = failure AND SENDER = $Agent_i$)
$FC_{ij} \leftarrow 0$; $Consider\_Help_j \leftarrow Consider\_Help_j + 1$; $Consider\_Rel_j \leftarrow Consider\_Rel_j + 1$;
If (PERFORMATIVE = inform AND SENDER = $Agent_i$)
$FC_{ij} \leftarrow 1$; $Consider\_Help_j \leftarrow Consider\_Help_j + 1$; $Consider\_Rel_j \leftarrow Consider\_Rel_j + 1$;
If (PERFORMATIVE = not_understood AND SENDER = $Agent_i$)
$VR_{ij} \leftarrow 0$;
If (PERFORMATIVE = not_understood AND RECEIVER = $Agent_i$)
$EC_{ij} \leftarrow 0$; $Consider\_Help_j \leftarrow Consider\_Help_j + 1$;
If (PERFORMATIVE = cancel AND RECEIVER = $Agent_i$)
$VR_{ij} \leftarrow 0$;

Figure 4: Measuring helpfulness and reliability in Request Interaction Protocol.

5 Reasoning About Reliability and Helpfulness

As discussed above, the previous analysis is based on externally measurable features of the interaction between the agents. However, in order to determine whether to accept a requested commitment or not, each agent needs to have some deliberative process, which may take into account the agent's current measurements of helpfulness and reliability.

As indicated previously there is an inherent conflict between being helpful and being reliable. The more helpful an agent is in making commitments, the greater likelihood there is that its ability and resources will not suffice and it will fail to live up to some of those commitments, thus being unreliable.

An agent will need to have some strategy regarding what commitments it makes (i.e. how helpful it is), and may also have some strategy regarding which commitments it should break in the event it finds itself unable to meet all its commitments. These strategies may be more or less "intelligent". They may also be either fixed, or dynamic, changing as the environment changes. In addition the agent may be self-aware and use this awareness to modify its strategies.

We outline briefly some ways in which the agent may reason about what to commit to, in order to achieve high reliability and/or high helpfulness ratings, or a high combined rating. Many more reasoning strategies can be defined, and the intent of this section is simply to provide some insight into how an agent can reason about ways to achieve greater responsibility, measured in the ways that we have defined.

5.1 Commitments according to capability

It has previously been argued by Padgham and Lambrix [PL00], in the context of a single agent making internal commitments (i.e. deciding which goals to pursue), that it seems unreasonable for an agent to commit itself to a goal that it is beyond its capability to achieve. For example it does not make sense for an agent to commit to a goal to cause rain, as it has no mechanism to achieve this goal. It would seem reasonable to require a responsible agent to also be constrained by its capabilities with regard to the social commitments which it makes.

However there is some disagreement in the literature as to how capability (or ability) should be defined. Some authors (e.g. [Sin99, vdHvLM94]) argue that an agent should be regarded as having the capability of achieving a goal only if there is a course of action available to the agent in the current environment. Thus capability is linked to current opportunity, as well as to general abilities. Padgham and Lambrix argue that adoption of goals should not be constrained by current opportunity, and thus have a looser definition of capability, requiring only that the agent has some knowledge of how to achieve the goal in some environmental situation. We refer to the former notion as *immediate capability* and the latter as *expected capability*. Immediate capability can be thought of as requiring that there be a a currently applicable plan (i.e. a plan instance where the preconditions are currently true). Expected capability requires that there is a plan in the plan library, which addresses the goal. More generally, expected capability can include the ability of the agent to generate (or otherwise obtain) a plan which it expects to be able to use.

In a conservative approach towards reliability we may wish to require that the agent makes a commitment only if it has an immediate capability with regard to the request. An approach which involves greater risk of being unable to fulfill the commitment, but which is more helpful, would be to make commitments whenever there is an expected capability.

5.2 Conflicts with previous commitments

Another factor affecting the ability of an agent to fulfill its commitment is whether or not the commitment conflicts with other commitments already made. For example an agent who is already committed to providing a particular an on-line booking facility for a particular football match with a specific quality of service may not be able to commit itself in a similar manner to another event in the same time frame.

Conflicts in general are not easy to detect in a generic manner. Thangarajah et al. [TWPF02] have done some work regarding detection and reasoning about resource conflicts, which are one of the common sources of conflict. They describe several levels of conflict, based on both the level of certainty that conflict will eventuate, and whether the conflict is limited to a specific time period, or whether, if it eventuates, it is ongoing. If the resource causing the conflict is reusable, then the conflict will resolve after one of the goals has finished using that resource. If the resource is such that once used it is permanently gone, and two goals both need it, then other methods are required to resolve the conflict.

By building on this or other work on conflict detection, it would be possible to have the agent consider definite or potential conflicts when making its commitments. As with capability, one can take either a more conservative or a more audacious approach towards conflicts under consideration. An agent could refuse all commitments with any type of potential conflict, or it could refuse only commitments resulting in certain conflict. The various options in between can be identified from the details of the work by Thangarajah et al. [TWPF02].

5.3 Strategies based on environmental conditions

In a dynamic environment it may be appropriate for the agent to adjust its commitment strategies depending on the characteristics of the environment. If there are relatively few requests, it may make sense to make commits more freely, as the opportunity to be helpful does not often arise. Conversely, when there are many requests and the agent is unable to meet them all, it may make sense to be more careful about commitments, perhaps taking a more conservative (or cautious) approach towards capability and/or conflict. In addition, if the environment is quite demanding, it may make sense for an agent to refuse a very resource intensive request as this would cause it to become unavailable for a period of time, thus reducing its ability to respond to less intensive requests, and thus decreasing its helpfulness measure.

An agent that is able to monitor and reason about the relevant environmental factors could choose its commitment strategy based on these, and would therefore behave differently in different domains. Such ability would also potentially allow the agent to switch strategies as the environment changed, to better adapt to changing circumstances.

It is also possible that in different application domains there are different preferences regarding reliability and helpfulness. For example, reliability is more important than helpfulness for the manager in a contract net protocol which is allocating specific tasks in building a house. In contrast, in a call centre, an agent working at a reception desk would be expected to have high helpfulness rating in dealing with all incoming calls.

These preferences could also affect the agent's reasoning about its commitments, in order to achieve the best behaviour for the specific domain.

5.4 Commitment with Self-awareness

An agent may also monitor its own behaviour in order to maintain reliability or helpfulness at a particular level, or in order to balance both to achieve a better responsibility rating.

If an agent is monitoring its own reliability and helpfulness scores, it may, for example, be cautious about making commitments (i.e. be less helpful), if its reliability is below a desired threshold, (thus negatively impacting the combined responsibility measure). Similarly an awareness that it has a lower helpfulness than reliability score may cause the agent to take more risks in order to positively affect its helpfulness. Similarly, awareness of its own performance on the combined responsibility measure could cause an agent to break a commitment to an already accepted request if that commitment conflicts with several new requests which it believes it would be more socially advantageous to accept.

This self awareness and its effect on the agent's strategies, can of course be combined with reasoning about environmental conditions. An agent that is able to change its strategy, based on both environmental conditions, and on an awareness of its own behaviour, is likely to be more successful in maintaining a high responsibility score.

6 Related Work

Our work has been inspired by related works in trust models [RHJ04, CF00]. In a multi-agent society, an agent always needs to decide when, how, and with whom to interact. Trust models, which normally gather knowledge about other agents, have been widely used in this decision making process. In order to computationally model an agent's trust in its opponents, numerical ratings are usually used for characterizing different levels of performance. There are different approaches regarding the retrieval and aggregation of ratings from different agents. For example, Yu and Singh use referrals to retrieval ratings [YS00, YS02]. A social network and corresponding techniques have been provided to gather information. Their initial work is based on the assumption that all witnesses are totally trustworthy, which may not be realistic in all circumstances. Hence, they proposed a model later for detecting deceptive agents in reputation management [YS03]. Schillo et al. [SFR00] also challenge the "'benevolence assumption" above by proposing a model to deal with lying witnesses. They have taken into account the probability of lying witness in their algorithm for trust, which enables agents to autonomously deal with deception and identify trustworthy parties in open systems.

Sabater and Sierra [SS01] proposed a more realistic approach by using reputation value. The reputation of an agent is determined by a combination of individual and group impressions. They discussed how impressions are formed, recorded, and also used in both individual and social dimensions.

In the domain of E-business, trust is used to build up reputations of members in a community. For example, Ebay (*www.ebay.com*) asks buyers to give feedback ratings after each transaction. Ratings from different buyers are summed together to get a seller's customer service reputation.

Our work differs from the above model in that our trust model is built without requiring ratings from other agents. The evaluation of an agent's trust can be based on its helpfulness and reliability, which both could be measured externally by counting and calculating the appearances of specific performatives, which are used in agent interaction. There are no further requirements, such as making agents responsible for giving ratings. To this end, our measure is an objective one, based purely on monitoring the agents' interaction and recording the number of relevant performatives.

The simplicity of this approach can also be a limitation, as we assume that it is straightforward to determine the commitments between agents purely from the messages exchanged. This will not apply to all protocols, particularly those in more homogeneous environments (as there is more implicit understanding between agents). We also assume that all commitments are unconditional, which greatly simplifies the determination of when a commitment is made.

Helpfulness and reliability highlight how socially responsible an agent is, and there has been a lot of previous work on these concepts. One of the key outcomes of this paper is to link these two concepts in a concrete way to the notion of social responsibility.

Social awareness is another aspect of social responsibility. A resource-bounded agent embedded in a dynamic domain should not only be able to reason about the changes of the environments on itself, but also be aware of the possible effects of its actions on the others in order to gain the optimal long term benefits from its community. That is, social awareness should be added into the individual autonomous agents to enable them to become more socially responsible. It is important for agents to be concerned with not only their own individual benefits but also those of society [JK95]. It has been shown by experiments that asking for and giving help are the most successful interaction attitudes [ACR96]. Gauthier [Gau75] even argues that individual utility maximisation destroys any real possibility of society. Helpfulness characterizes that an agent honors its social role and tends to give help to the others in need.

However, with regard to giving help, Castelfranchi[Cas95] believes that an agent is not rational if it always adopts other goals on request, where there is no benefit for the agent by achieving the adopted goals. Newell also states the principle of rationality as "If the agent has knowledge that one of its actions will lead to one of its goals, then the agent will select the action." [New82]. Hence, a "benevolent agent" or an "over-helpful agent", which always adopts other's requests as its own goals, is irrational [Ld96, JC97]. In contrast to this black-and-white approach, we have defined the helpfulness of an agent to be a value between 0 and 1 which is based on measured realities whilst still allowing the agent to determine whether or not it will be helpful.

Intelligent agents normally should have self-awareness and also awareness of its embedded environment (at least partially). McCarthy gives a detailed discussion about what conscious machines should have, and also some mechanisms for achieving this [McC95, McC04]. In our work, this corresponds to an agent adjusting its behaviour according to its social responsibility ranking, using its awareness of both the environment and itself as inputs.

7 Conclusion

We have shown how social characteristics such as reliability and helpfulness can be measured in an objective way, and how this can be used as a measure of social responsibility. In particular, we have shown how approaches can use particular specifications such as the FIPA ACL and Interaction Protocols.

This provides an initial attempt to measure in an objective manner, important social characteristics. It was noted that some protocol definitions do not facilitate recognition of commitments, and thus cannot be monitored for social characteristics based on such notions. However relatively minor adjustments to the protocols would support such monitoring. A recommendation from this work would be that an important characteristic to consider when defining protocols, is the ability to monitor externally for an agent's social behavior with regards to the protocol.

We strongly believe that much of the value of a notion of social commitment lies in the ability to observe the behaviour surrounding that commitment. One item of further work is to perform some experimental validation of this approach, using a variety of protocols.

As mentioned above, one limitation of our work is that we assume that a commitment can be determined purely by whether a particular message has been sent. For some more complex protocols, or for conditional commitments, a deeper analysis of the interaction may be required, such as tracking the status of conditional commitments through the execution of the protocol, or searching through the execution sequence for a particular sequence of messages. Whilst such extensions will complicate the issues of pending commitments, it will allow a greater range of protocols to be used without abandoning the objective (and verifiable) measurement of reliability and helpfulness.

References

[ACR96] M. Miceli A. Cesta and P. Rizzo. Effects of different interaction attitudes on a multiagent system performance. In W. van de Velde and J. Perram, editors, *Agents Breaking Away*, pages 128–138. Springer-Verlag, 1996.

[ARS+05] Ronald Ashri, Sarvapali Ramchurn, Jordi Sabater, Michael Luck, and Nicholas Jennings. Trust evaluation through relationship analysis. In *Proceedings of the Fourth International Conference on Autonomous Agents and Multi-Agent Systems*, Utrecht, July 2005.

[Cas95] C. Castelfranchi. Guarantee for autonomy in cognitive agent architecture. In *Proceedings of ECAI-94 Workshop on Agent Theories, Architecture, and Language.* Springer-Verlag, Berlin, Germany, 1995.

[CF00] C. Castelfranchi and R. Falcone. Social trust: A cognitive approach. In C.Casterfranchi and Y.H.Tan, editors, *Trust and deception in virtual societies*, pages 55–90. Kluwer Academic Publisher, 2000.

[CMK66] C. W. Cleverdon, J. Mills, and E. M. Keen. *Factors Determining the Performance of Indexing System.* Aslib Cranfield Research Project, Cranfield, UK, 1966.

[FC02] N. Fornara and M. Colombetti. Operational specification of a commitment-based agent communication language. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-agent Systems*, pages 536–542, 2002.

[FC04] Nicoletta Fornara and Marco Colombetti. A commitment-based approach to agent communication. *Applied Artificial Intelligence*, 18(9-10):853–866, 2004.

[fIPAFa] Foundation for Intelligent Physical Agents (FIPA). Communicative act library specification. Available at standards.computer.org/fipa/specs/fipa00037.

[fIPAFb] Foundation for Intelligent Physical Agents (FIPA). FIPA request interaction protocol specification. Available at standards.computer.org/fipa/specs/fipa00026.

[Gau75] D. Gauthier. Reason and maximisation. *Canadian Journal of Philosophy*, 1975.

[HJ97] L. M. Hogg and N. R. Jennings. Socially rational agents. In *Proceedings of AAAI Fall Sym. on Social Intelligent Agents*, pages 61–63. Boston, USA, 1997.

[JC97] N. R. Jennings and J. R. Campos. Towards a social level characterisation of socially responsible agents. In *IEE Proceedings on Software Engineering 144 (1)*, pages 11–25, 1997.

[JK95] N. R. Jennings and S. Kalenka. On social attitudes: A preliminary thoughts. In *International Workshop on Decentralized Intelligent and Multi-Agent System*. Krakow, Poland, 1995.

[Ld96] M. Luck and M. d'Inverno. Engagement and cooperation in motivated agent modelling. In Zhang and Lukose, editors, *Distributed Artificial Intelligence Architecture and Modelling: Proceedings of the First Australian Workshop on Distributed Artificial Intelligence*, pages 70–84. Springer-Verlag: Heidelberg, Germany, 1996.

[McC95] J. McCarthy. Making robots conscious of their mental states. In *Working Notes of the AAAI Spring Symposium on Representing Mental States and Mechanisms*, Menlo Park, CA, 1995.

[McC04] J. McCarthy. Notes on self-awareness. Available from www-formal.stanford.edu/jmc/selfaware.html., 2004.

[New82] A. Newell. The knowledge level. In *Artificial Intlligence*, pages 87–127, 1982.

[PL00] L. Padgham and P. Lambrix. Agent capabilities: Extending BDI theory. In *Proceedings of Seventeenth National Conference on Artificial Intelligence*, pages 68–73, 2000.

[RBJ89] V. Raghavan, P. Bollman, and G. Jung. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems*, 7(3):205–229, 1989.

[RG91] A. Rao and M. Georgeff. Modeling rational agents within a BDI-Architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, 473-484*, 1991.

[RHJ04] S. D. Ramchurn, D. Huynh, and N. R. Jennings. Trust in multi-agent systems. In *The Knowledge Engineer Review*, 2004.

[SFR00] M. Schillo, P. Funk, and M. Rovatsos. Using trust for detecting deceitful agents in artificial societies. In *Applied Artificial Intelligence*, pages 825–848. Taylor and Francis, 2000.

[Sin91] M. P. Singh. Social and psychological commitments in multiagent systems. In *In AAAI Fall Symposium on Knowledge and Action at Social and Organizational Levels*, pages 104–106, 1991.

[Sin99] M. P. Singh. Know-how. In *Foundations of Rational Agency*, pages 105–132. Kluwer, 1999.

[Smi80] R. Smith. The contract net protocol. *IEEE Transactions on Computers*, 29(12):1104–1113, 1980.

[SS01] J. Sabater and C. Sierra. REGRET: reputation in gregarious societies. In Jörg P. Müller, Elisabeth Andre, Sandip Sen, and Claude Frasson, editors, *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 194–195, Montreal, Canada, 2001. ACM Press.

[SS02] Jordi Sabater and Carlos Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the First International Conference on Autonomous Agents and Multi-Agent Systems*, pages 475–482, Bologna, July 2002.

[TWPF02] J. Thangarajah, M. Winikoff, L. Padgham, and K. Fischer. Avoiding resource conflicts in intelligent agents. In *Proceedings of the 15th European Conference on Artificial Intelligence*, 2002.

[vdHvLM94] W. van der Hoek, B. van Linder, and J. Ch. Meyer. A logic of capabilities. In *Logical Foundations of Computer Science*, pages 366–378, 1994.

[VR79] C. J. V. Rijsbergen. *Information Retrieval, 2nd edition*. Dept. of Computer Science, University of Glasgow, 1979.

[YS00] B. Yu and M. P. Singh. A social mechanism of reputation management in electronic communities. In *Cooperative Information Agents*, pages 154–165, 2000.

[YS01] P. Yolum and M. P. Singh. Commitment machines. In J. Meyer and M. Tambe, editors, *Proceedings of the Eighth International Workshop on Agent Theories, Architectures, and Languages (ATAL-2001)*, pages 245–257, 2001.

[YS02] B. Yu and M. Singh. Search in referral networks. In Bin Yu and Munindar P. Singh, editors, *Proceedings of the AAMAS Workshop on Regulated Agent-Based Social Systems: Theories and Applications*, 2002.

[YS03] B. Yu and M. Singh. Detecting deception in reputation management. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 73–80, 2003.