Trust Attributes, Methods, and Uses
An Extended Abstract

Gabriel Becerra and Jason Heard and Rob Kremer
Department of Computer Science
University of Calgary
Calgary, AB, Canada
{ayala, heard, kremer}@cpsc.ucalgary.ca

Introduction
Multi agent systems (MAS) are a collection of multiple, possibly independent entities (called agents) working in a common environment. One goal of MAS studies is to achieve synergy, or to achieve more through cooperation or coordination between agents than would otherwise be possible (Denzinger 1995). Open systems are one type of MAS in which agents from untrusted sources may operate (Hewitt 1985). Within an open system, there must exist either some form of control over the behaviour of individual agents or some method for agents to evaluate which other agents can be trusted. The issue of trust itself is not new, however, and has been discussed extensively by many researchers in psychology, management, communications, sociology, economics, and political sciences well before trust because useful to the study of agents (McKnight & Chervany 2001).

Up to the year 2001 there were at least 65 cited articles and monographs defining the concept of trust. Out of those 65 articles 23 came from psychology; 23 came from management and communications; and 19 came from combined areas such as sociology, economics, and political sciences (McKnight & Chervany 2001). This large number of definitions has contributed to the ambiguity that surrounds the concept of trust.

Mayer et al set the groundwork for the next generation of trust research and its goal: define trust models that accommodate a specific definition of trust (Mayer, Davis, & Schoorman 1995). As a response to this call to model, researchers begun constructing models based on the properties they considered important or relevant to their context, be it sociology or economics. While this work succeeded in creating several models helpful to several different specific situations, it failed to generate models that were neutral.¹

This paper will attempt to distill the many definitions and models of trust into several key portions similar to the way that Jennings et al aided the discussion of agents in general by outlining a set of possible descriptors for agents, including reactive, social, and rational Jennings, Sycara, & Wooldridge (1998). Once completed, this paper will then use these new terms to describe several models of trust.

The Terms of Trust
If Alice trusts Bob to perform some task, she must have determined that his attributes meet some set of requirements. This paper will divide the trust terminology into three categories. The first is the set of attributes of Bob that she may consider. The second is a set of methods of discovery that Alice can use to determine those attributes. The third is a set of methods of evaluation Alice will use to evaluate the attributes as she has determined them.

This section will outline each of the terms that are used repeatedly in the area of trust, and will list a selection of authors that have used these terms (or synonymous terms to date).

Attributes
Four main attributes must be considered to evaluate whether Bob can be trusted to perform the required task. The first is his integrity, the second is his motivation, the third is his predictability and the fourth is his competence. Integrity defines the level of trustworthiness of Bob, while motivation defines how motivated he is to perform the given task. Predictability describes how regular Bob's actions are. Finally, competence describes Bob's ability to perform the required task.

Integrity
The attribute of integrity describes how ethical Bob is in general. This can describe how ethical, honest or moral an agent is. This can also be called Bob’s willingness to act (and behave) properly. Instances of this attribute are found in the following papers:

Hartman (1999) considers this property as a type of trust whereas in this work we have classified it as an attribute which Alice seeks to determine in order to measure the trustworthiness of Bob.

Mayer, Davis, & Schoorman (1995) consider that the relationship between integrity and trust involves Alice’s perception that Bob adheres to a set of principles that Alice finds acceptable.

McKnight & Chervany (2001) consider that when Alice securely believes that Bob makes good faith agreements with other agents, tells the truth, and fulfills its promises,
Bob is an ethical agent that truly cares about Alice’s interests. This property is attached to a belief condition that leads towards the proper foundation of intention.

**Motivation** An area that is less well studied includes the motivation of Bob to complete the task. This may be an area of future study, because as Bob’s motivation increases, Alice relies less on Bob’s integrity.

**Predictability** Both prediction and trust are means of uncertainty reduction. To be meaningful, trust must go beyond predictability. In other words, Bob’s predictability is insufficient to make Alice take a risk and put herself in a vulnerable situation. It is, however, helpful for Alice to determine the predictability of Bob. Instances of this property are found in the following papers:

Mayer, Davis, & Schoorman (1995) consider predictability as a factor that influences the cooperation between two agents. If Alice expects that Bob will predictably behave positively, Alice will have the disposition to cooperate with Bob.

McKnight & Chervany (2001) argue that Alice is willing to depend on, or intends to depend on, Bob in a given task or situation with a “feeling of relative security”, even though negative consequences are possible. This property is composed of five elements: (1) Alice may have to deal with negative consequences or risk in unfamiliar or uncertain situations, (2) Alice’s readiness to depend or rely on Bob is central to trusting intentions, (3) Alice feels safe, assured, and comfortable about the prospect of depending on Bob, (4) trusting intention is situation and person specific, and (5) trusting intention involves willingness, which is not the same as having control over Bob.

van Witteloostuijn (2003) explores the conditions under which trustworthy behaviour can be expected. Witteloostuijn study is based game theory and aims to the comprehension of how trust and cooperation are closely related. Moreover, Witteloostuijn considers that cooperation is a consequence of the unobservable cognition of trust.

**Competence** Competence is an attribute that describes how competent Bob is for the task that Alice wishes him to perform. It is possible that Bob only has a specific competence, which indicates that Bob can only perform the selected task in a certain set of situations ($Sit' \subset Sit$). Instances of this attribute are found in the following papers:

Falcone & Castelfranchi (2001) argue that this attribute is present when Alice believes that there exists an agent, say Bob, that has the power to achieve the task. The authors differentiate between two types of delegation: weak and strong. The first type refers to cases where there is no explicit delegation of the task from Alice to Bob. The specific task is accomplished by the exploitation of an autonomous action of Bob. The second type of delegation exists when Alice explicitly delegates the task to Bob. In this case, there is a willingness to depend on Bob’s competence in such area.

Hartman (1999) identifies competence as another type of trust. As with integrity, we classify competence as an attribute and not as a type of trust. Hartman considers that this property allows Alice to assess the extent to which its interests will be cared for and be protected. Moreover, this question will also help evaluating the consistency of Bob’s behaviour while performing the task.

Mcknight & Chervany (2001) attach competence to a belief condition that serves as the foundation of the intention property.

Mayer, Davis, & Schoorman (1995) developed the competence property under the term “ability”. Ability is that group of skills, competencies, and characteristics that enable Bob to have influence within some specific domain. The domain of the ability is specific because Bob may be highly competent in some technical area. However, such competence may be limited by the context or the specificity of the task.

**Methods of Discovery**

There are four main methods that can be used to discover Bob’s attributes. They are intuition, experience, hearsay, and records. Intuition includes all methods that Alice can use to determine Bob’s attributes without considering Bob specifically. Generally this focuses on how agents in general would react to the current situation. Experience focuses on personal experience with Bob. Hearsay is the opinions of other agents who are not necessarily trusted. Records are any institutionally supported data about Bob.

**Intuition** The concept of intuition is hard to adapt to multi agent systems. It is hard to think of the “gut” feelings that Alice may have towards Bob. However, to clearly define the meaning of this method, we will define an intuition as any computation that Alice can perform before her first conversations with Bob. This can include game theoretic approaches to the situation not specific to the agent Bob. In general, this method can help Alice determine Bob’s motivation to complete the task. Instances of this property are found in the following papers:

Romah & Hartman (1999) consider this as their third type of trust in their model. As with their previous classifications, we have to disagree with such classification. However, we agree on the description of the attribute and its subdivision (rational and irrational intuition). The description of this question can be summarized as an aid for Alice when measuring how right a situation, deal, or relationship with Bob feels. This question may lead to complications while implementing an agent-based system if we consider that intuition depends too much on the so-called gut-feelings. The latter is what the authors considered irrational intuition or the chemistry between two parties. On the other hand, the rational part of this property can be considered in an agent-based system by applying some heuristics to Bob’s competence level.

Denzinger & Hamdan (2004) investigate tentative stereotype models based on agents’ situation-action pairs. The results of this investigation allows agents to predict an agent’s behaviour even after few observations. Their models can be reevaluated in order to switch from the original impression to other type of stereotype.
Experience Alice, as a cooperative agent, may have had previous experiences with Bob. In addition, Alice may have observed Bob as he worked with another agent. These interactions may have been all related to the actual task or another set of tasks that Bob may also be capable of doing. This method can help Alice assess Bob’s integrity and competence. Instances of this property are found in the following papers:

Ramchurn et al. (2004) make use of this method in order to assess trustworthiness under the concept of confidence, which they describe as a measure of certainty, based on evidence from past direct interactions with Bob.

Bower, Garber, & Watson (1997) suggest and analyze uncertainty and learning about the populations of agents. The analysis shows how the degrees of trust by a principal, Alice, and cooperation by another agent, say Bob, can depend on the past behaviour of that other agent. That is, such analysis depends on the experience Alice has had with Bob.

Hearsay As Alice communicates with more agents, she may be able to obtain data on Bob’s integrity and competence through third-party sources, without direct interaction (Mayer, Davis, & Schoorman 1995). Instances of this property are found in the following papers:

Falcone & Castelfranchi (2001) deal with this matter by applying transitivity concepts. Let $\prec$ mean trust and $George \in Ag$. If $Alice \prec Bob$, and $Bob \prec George$, then $Alice \prec George$.

Ramchurn et al. (2004) make use of this method in order to assess the trustworthiness under the concept of reputation, which is described as Alice’s measure of certainty (leading to trust), based on the aggregation of confidence measures provided to it by other agents that have previously interacted with Bob.

McKnight & Chervany (2001) treat this property under the concept of influence by stating that when Alice allows Bob to influence it, Alice is depending on that opinion to be correct given the fact that bad consequences may follow if it is misdirected on another party’s opinion.

Records Another method of obtaining information about Bob is through records from an institution or authority. In this case, Alice still obtains information about Bob, but the this party is an agent that represents an institution or an authority.

While Carter and Ghorbani (Carter & Ghorbani 2004) consider the case where Alice should be aware of the trust level she has towards an institution, and in this way, records could be considered a special case of hearsay, we believe that this method deserves its own separate method.

Methods of Evaluation

Once the attributes of Bob have been discovered as much as is possible, they must then be evaluated to determine whether Bob can be trusted. To do this, Alice must consider two factors. How much does Alice risk to trust Bob, and how willing is Alice to take that risk.

Level of Risk Risk is considered to be an essential component of trust. It is important for researchers to understand that risk is inherent in the behavioural manifestation of the willingness to be vulnerable. Trust will lead to risk taking in a relationship, and the form of the risk taking depends on the situation.

We emphasize the fact that vulnerability is a consequence of taking the risk of trusting other agents. Instances of this property are found in the following papers:

Gambetta (2000) considers trust as a particular level of the subjective probability with which Alice assesses that Bob will perform an action. For trust to be present there must be the possibility for disappointment or betrayal.

Mayer, Davis, & Schoorman (1995) provide one of the most used definitions of trust based vulnerability, which is defined as “the mutual confidence that no party to an exchange will exploit another’s vulnerabilities”.

Ford (2003) while investigating the concept of knowledge sharing took into consideration this particular question in the sense that trusting other parties with valuable information involves a high-level risk.

Bower, Garber, & Watson (1997) develop the concept of risk within the area of game theory. They consider that ”trusting is risky for Alice because there are two types of agents and Alice is uncertain about the type it is confronting. When a rational (opportunistic) agent is allowed discretion, its immediate payoff is higher if it chooses to betray the trust of Alice, but Alice’s immediate payoff is higher if Bob chooses to cooperate”.

Willingness to Trust According to Mayer et al (1995), the disposition to trust is an internal factor that will affect the likelihood of Alice to trust other agents. Such a disposition to trust might be thought of as the general willingness to trust others.

Given a situation and the internal data of Alice, she will have a disposition or willingness to either depend on Bob or perform certain action towards a specified task (Mayer, Davis, & Schoorman 1995). Based on this duality, we identify the following subdivision: the willingness to depend on another agent, and the willingness to act or achieve a specified task. Instances of this property are found in the following papers:

McKnight & Chervany (2001) This question is linked to a belief condition that leads toward the proper foundation of intention.

Ford (2003) by using the term behavioural trust is able identify that the central question is whether Alice is willing to depend on Bob or not. The willingness of Alice to be vulnerable to Bob’s actions, are based on the expectation that Bob will perform a particular action important to Alice, irrespective to the ability to monitor or control Bob. The same definition is used by Carter and Ghorbani (Carter & Ghorbani 2004) while elaborating a new trust model based on reputation.

Hartman (1999) relies on the existence of legal systems in the absence of integrity. This legal system will decrease Alice’s risk.
Case Studies

This section will show how the terms defined above can be helpful in describing and comparing several trust systems.

The ART Testbed

The Agent Reputation and Trust (ART) Testbed is a multi agent game designed to be a testbed for trust issues (Fullam et al. 2005). The testbed simulates an open multi agent system where art appraisal agents communicate with the purpose of increasing the accuracy of their appraisals. Each competitor in the system controls one art appraisal agent. Each agent is assigned capabilities for appraising various categories of art work. The system provides clients who will make appraisal requests of each agent. When an agent doesn’t possess the ability to appraise a piece of artwork, that agent can request an opinion from another appraisal agent. The other agent has the option of being dishonest in both the expression of its ability to appraise the artwork and in the actual opinion sent back to the first agent. To aid the agent in determining which agents are trustworthy, they can ask other agents for their evaluation of the other agents’ reputations.

A successful agent in the ART testbed must take both integrity and competence into account. Integrity comes from the knowledge of how well other agents self-report their capabilities when they have been queried for an opinion. Competence is important since each agent has its own level of knowledge for various types of artwork. Predictability is also a factor, as it relates to competence. An agent in the ART framework is capable of using intuition, experience and hearsay. Intuition could be used to consider how agents’ behaviours might change over the course of the game. Experience and hearsay are the main mechanisms supported and encouraged by the testbed’s design. The exclusion of reports from an authority is probably necessary to maintain the focus on reputations and opinions.

CASA (Cooperative Agent System Architecture) and Social Commitments

This section will introduce and evaluate the CASA system as a basis of trust and agents (Kremer, Flores, & Fournie 2004). CASA is a communication-based multi agent system written in Java. CASA has several unique features that could be useful in the exploration of social commitments and trust.

A cooperation domain (CD) is an agent within CASA designed to aid the communication of many agents in large groups. The CD allows agents to communicate with one another without knowing about every other agent. This is similar to how someone must only register with a listserv to receive all communications sent by members of a group, and to send an e-mail to one address in order for everyone in the group to receive it. Agents register with the CD, and as a result they receive all non-private communications that are sent to the CD (including those they send). Figure ?? shows four agents participating in a conversation through a cooperation domain. The double-headed arrows indicate two-way communication between agents.² CASA agents are free to communicate with each other directly (they can bypass the CD), but they will potentially lose the use of services offered through the CD and, relating to trust, the service of fingering or sanctioning anti-social agents.

CASA provides a useful tool for researching social commitments and trust: the CASA system implements several conversation policies. These policies are used by CASA agents to define the communication protocols used. Conversation policies map messages with certain performatives to the creation and resolution of social commitments. Kremer and Flores describe how the CASA system defines social commitments and bases the conversation policies on the performatives in CASA from FIPA (Kremer & Flores 2005).

Because the communication system in CASA is based upon social commitments, it is easy for an agent to monitor communications of itself (or in some special cases, all agents in a CD) and determine which agents are breaking social commitments. Agents who monitor their own communication can determine other agents integrity and predictability based on personal experience. In addition, if an agent communicates within a cooperation domain, it is possible to include a system sanctioning agent which removes agents that pass some level of anti-social behaviour—this may increase an agent’s willingness to trust agents within he system (Heard 2005). Another option is a BBB (better business bureau) agent which will, upon request, provide records on the integrity of agents within that cooperation domain. These services are provided by trusted agents which are privy to all communications through the cooperation domain.

At this point, the areas that CASA does not provide assistance with are competence, motivation, intuition and hearsay. These areas could be implemented in individual agents, but have not been at this point.

Adaptive Trust and Cooperation: An Agent-Based Simulation Approach

This project focuses on applying agent-based technologies both to simulate inter-firm relations and to determine how cooperation, trust and loyalty emerge from agent-based computational economics models instead of following a transaction cost economics model (Nooteboom, Klos, & Jorna 2001). Briefly, this market-driven project aimed at identifying how trust and loyalty affect each other under several circumstances.

In this model, the attributes described above (e.g. integrity and competence) are determined by applying the following methods of discovery: intuition (rational), personal experience, and hearsay. Rational intuition is used by applying an economic selection on a set of possible partners. The agent asks itself: how profitable will this partnership be? In addition to this internal computation, the agent also gathers information from its previous interactions with such agent. Externally, the agent will gather information from transaction partners. This because it is cheaper to gather information from other sources than to be extremely cautious in events where cooperation from both sides will increase profits. Consequently, if profits are increased, then trust andloy-pants (broadcast), to a specific subset of the participants (multi-cast), or to a single agent (directed).
alties toward each other will also increase. The latter brings up the concept of predictability since agents in this simulation wonder about what the behaviour of their partner would be in cases where the partnership has not been as effective as desired. Finally, after applying the methods of discovery and determining the proper values for the attributes, the agent then decides on its willingness to cooperate. This is reflected by its risk factor, which affects its behaviour given certain circumstances. In other words, the agent computes values that serve as thresholds that indicate the best moment to defect and terminate its commitments.

The model used by Nooteboom et al does not present any reference to methods of discovery such as records, motivation, and did not specify how hearsay exactly affect the trust level of one agent toward another agent. It is important to mention that the authors never make use of integrity, but it is easy to determine that an agent’s integrity in this evaluation is not very important. This because agents, using this model, defect as soon as the profits or other offers are close to their thresholds (either lower or upper bounds).

Conclusion

Discussing trust requires a well-defined set of terms for trust. As researchers have discovered, redefining trust for each application is not always helpful. This paper hopes to aid in the discussion of trust by providing a set of descriptive terms that can be used to describe the areas of trust that a particular model or implementation uses. The terms provided in this paper will hopefully allow researchers to easily identify the attributes and methods being used in various models and projects.

References


