

Mediated agent interaction

Enrico Coiera
Centre for Health Informatics
University of New South Wales,
UNSW NSW 2055, Australia
ewc@pobox.com

Abstract: This paper presents a framework for agent communication and its mediation by technological systems. The goal of the framework is to provide quantitative mechanisms that will allow principled decisions to be made about the use and construction of mediating technological systems. Beginning with a simple model of interaction between agents, a model of communication influenced by bounded knowledge of others (or common ground) is developed. This leads to predictions that agent interactions generate equilibrium phenomena where optimal levels of grounding emerge over time between agents.

Introduction

Although computational and communication systems have had a largely independent development over the last century, there is no sharp demarcation between them either at a theoretical or technical level. Computational systems are as dependent on the transmission of data during calculation, as communication systems are dependent on the encoding and decoding of data for transmission. Indeed, a complex system like the World Wide Web clearly functions only because of the way that it marries communication and computation.

While there are bodies of work that either examine the human use of media, the transmission characteristics of media, or the computation of messages, we lack an integrated account of all three. At present we have strong models for several of the components of mediated interaction. Information theory models the process of message transmission across a channel. Computational theories model how an agent, given certain input, internal goals and resources, can construct or respond to a message. However, we have no adequate theoretical account that explains the behaviours that emerge when agents and channels combine to create a mediated interaction system.

This paper reviews the different bodies of communication research, identifies their specific deficiencies, and then outlines the scope for developing a new mediated communication theory. A framework for such a theory is then developed, based upon the

notion of bounded mutual knowledge, or common ground, that is shared between communicating agents (Coiera, 2000). An examination of the costs of the grounding process then leads to the view that grounding is a kind of cost minimisation process, based upon maximisation of agent utility, and results in cost-minimising equilibria developing between agents.

Three Views of communication

Communication is the way we make our intentions known to the world, interact with others to achieve our goals, and the way that we learnt about our environment. Perhaps more than any other skill, the ability to hold conversation is the defining achievement of any intelligent agent. To do so requires knowledge of self and of other, and the ability to engage in a process of exploration whose bounds are imposed by mutual consent through interaction.

There are also physical bounds to communication, and these are provided both by the physical organs of agents, and by the communication channels they use to transmit messages. Communication channels are the intermediating links that agents use to converse, and the properties of channels also limit the forms a conversation can take. Sound, image and text all support very different conversational structures. Thus, communication is shaped both by the agents who wish to converse, and the mediating environment within which the conversation takes place.

Given its centrality in human society, communicative interaction has been looked at extensively over the centuries. Some have sought to develop technologies for communication or computation, whilst others have been motivated to understand and characterise the various forms of human interaction. Given these different intentions, we have communication theories that either emphasise the physical bounds to communication in technological systems, the knowledge bounds limiting rational agents, or the relativity of meaning arising out of the effects of context. It would be foolhardy to simplify these vast and complex bodies of work, but of necessity we can coarsely caricature three separate points of view:

- *Linguistic and social communication theories* (context centric) - Accounts of human communication have long been central to artists, musicians and writers, who have a profoundly different way of viewing communication to technologists. Fields like semiotics, communication theory (Cobley, 1996), and philosophy have sought to emphasise how the context of communication profoundly alters the meaning that is drawn from a message. These ideas have coalesced into the notion of *post-modernism*, which emphasises the relativity of meaning for individual agents. More recently linguists have emphasised that language use, and the structures of conversations, are actively shaped within the context of a discussion (Clarke, 1996). Whilst providing a rich body of ideas, context centric theories are essentially qualitative, and do not provides us with quantitative mechanisms for making decisions about the design or management of communication technologies or computational processes.
- *Information theory* (channel centric) - In the classic information theoretic setting described by Shannon, a sender and a receiver exchange messages across a channel.

The effect that different channels have in distorting the message is modelled statistically as the effect of noise. Information theory concerns itself with the most efficient way to encode a message given the space of all possible messages and the effects of noise on any given channel (van der Lubbe, 1997). However, information theory says nothing about the meaning of a message, or how agents structure a message to suit a given context.

- *Computational theories* (model centric) - The communicative behaviours of agents can be partially accommodated in computational theories. The intentions and knowledge of each agent, as well as its decision procedures for constructing messages, can all be explicitly modelled as computable actions (Russell and Norvig, 1995; Moulin and Chaib-Draa, 1996). In particular, interacting agents can be modelled as Turing machines (Herken, 1988). However, a Turing machine assumes messages are sent across a cost-free channel. Indeed, channels are usually not modelled at all in theoretical accounts of computation, but are left implicit. This assumption is often explicitly acknowledged as "cheap talk" in computationally inspired decision models like Game Theory (Bierman and Fernandez, 1995).

Recently, there has been some movement between these three views. For example, context centric theorists have begun to realise that they must move away from highly-context dependent research to explore more general communication phenomena (e.g. Berger, 1991). At the other end of the spectrum, designers of interactive computer technologies have begun to admit the need for some understanding of the context of communication in the human-computer interaction. (e.g. Nardi, 1996). In artificial intelligence circles, there has been a movement away from the model-centric view of computational intelligence towards one of *situated cognition* (Suchman, 1987). Here, the situation in which an agent interacts with the world in some way determines the behaviours that agents display (Brooks, 1991). Biological comparisons with the simple decision rules that drive organisms like ants are often used, pointing out the complex social behaviours that emerge through interaction with environmental structures.

The Scope and role of a mediated communication theory

Individually, the channel, context or model-centric views are limiting if we want to develop technologies that craft themselves around the meaning of conversations. If we wish to build communication systems optimised to the specific task needs of populations, that optimise their behaviour based upon the intent behind a conversation, or if we want to build autonomous software agents that have some scope for exploratory interaction with each other, then we need to close the gap between current theories. If successful, such a union of theories should allow us to answer two distinct classes of question:

- *Local communication decisions*: How should an agent communicate to maximise the outcome of an interaction? What combination of channels and media would be most favourable given the context of the interaction? Indeed, given a set of other agents to interact with, which ones are likely to be favourable choices, and what should be said to them to achieve the aim of the interaction? The resources available will always bound such choices for agents.
- *Global communication decisions*: Given an existing population of agents and communication technologies, what would the consequences be of introducing new

communication technologies, or varying existing ones? Could the impact of the Web have been predicted in advance? Although we might find it hard to predict what individual agents might do in any given circumstance, we could make useful predictions about groups of agents. For example, given an organisation with a set of information and communication problems, what mix of mediating technologies will provide the best solution?

The ultimate goal in answering such questions would be to identify quantitative mechanisms that allow principled decisions to be made about the use and construction of *mediated interaction systems*. By analogy, the periodic table arranged elements in a particular structure and led to the identification of previously unknown elements. Similarly, one could hope that an interaction framework would provide coherence to our observations on the use of media, predict how different interaction settings will function, and where feasible, identify new forms of interaction and interaction technologies.

A general setting for communicative interaction

Intuitively, an interaction occurs between two agents when one agent creates and then communicates a message to another. The intent of that act is to influence the receiving agent in some new way. A *mediated interaction* occurs when a communication channel intermediates between agents by bearing the messages between them. We will begin by developing a general model of what it means to have an interaction between two agents.

Definition 1: An **interaction** occurs between two independent agents A_1 and A_2 when A_1 constructs some message m , and passes it to A_2 across some communication channel c .

$$A_1 \xrightarrow[c]{m} A_2$$

The interaction between A_1 and A_2 occurs because A_1 has some intentions, and has decided to influence A_2 via the message, to satisfy its intention. We have thus placed our agents within a world in which they have choice over their own actions, and some ability to interact with the world to change it to suit their purposes.

In this model, agents are independent entities. Agents have no direct control over the behaviour of others (they can only attempt to influence behaviour through interaction), and they have no privileged knowledge of the internal state of other agents (they can only observe each other's behaviour in the world, and infer their internal state). Agents also operate with finite resources, and consequently need to minimise costs on resources and maximise the benefits of any action they contemplate.

The interaction between agents is dependent upon three things - the channel, the relationship between the agents, and the context within which they find themselves. We need to expand upon the simple interaction model presented above, to incorporate each of these elements. Of the three, most will be said about the effect of agent relationship upon communication.

Effects of channel upon interaction

Standard information theory describes how the outcome of an interaction is determined in part by the characteristics of the mediating channel c (van der Lubbe, 1997). A message m is transmitted between the agents as a data stream D . The communication channel c is not neutral to D , but has a modifying effect upon it. Data may be affected by noise in the channel, for example. So, we note that when agent A_1 utters a data stream D_1 , that it is then modified by c , to be received as D_2 by agent A_2 (Figure 1). Since the information theoretic aspects of channel noise on message transmission are standard results, they need not be described any further here, except to reinforce their importance in quantifying the overall characteristics of inter-agent interactions.

Effect of agent relationships upon interaction

Agents have independent origins, independent resources and independent prior experiences, which modify their internal state over time. Thus, in this setting we are not permitted to look inside an agent and know what it knows, only to infer that from observable behaviour. Critically then, the private world models of the communicating agents M_1 and M_2 are not verifiably identical to any observer.

We can now refine our interaction setting to include this new element. An individual agent A_1 creates a data stream D_1 , based upon its knowledge of the world M_1 . A_2 receives D_2 , and generates its own private interpretation of its meaning based upon D_2 (Figure 1).

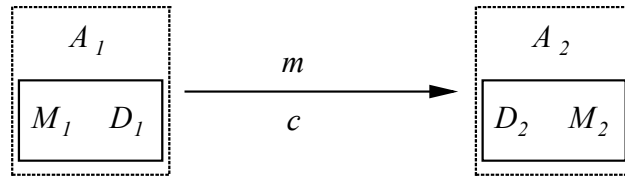


Figure 1: The M-model of interaction (after McCarthy and Monk (1994)).

We next note that agents communicate more easily with other agents that have similar experiences, beliefs and knowledge. Intuitively we know that it takes greater effort to explain something in a conversation to someone with whom we share less common background. Conversely, individuals who are particularly close can communicate complicated ideas in terse shorthand.

Agents who share similar background knowledge can be termed *homophilus*, and dissimilar agents *heterophilus* (Lazaarsfeld and Merton, 1964). The notion of homophily arose out of the analysis of social relationships between individuals of similar occupation and educational background. Homophilus individuals tended to have more effective communication amongst themselves, and the more individuals communicated, the more homophilus they became (Rogers, 1995).

The degree of agent similarity is thus based upon the similarity of their knowledge about the world:

Definition 2: For interacting agents a and b with respective internal models M_a and M_b , with respect to a given message m , the degree of **homophily** between them $H_{a:b}$ is the ratio $f(M_a)/f(M_b)$, where $f(x)$ is some well-defined function on M_x .

This definition has the property that identical agents would have unit homophily. Further, one would expect that the cost of communicating between dissimilar agents is a function of the degree of homophily:

When $H_{a:b} = 1$ the interaction is *relationship neutral*. The agents are able to communicate with no effort diverted to managing misunderstandings, other than those caused by the effects of channel, such as noise.

When $H_{a:b} > 1$ the interaction is *relationship negative*. This means that agent a needs to explain part of a message to b , by transmitting background knowledge along with the principle message to permit b to fully understand. The negativity is associated with the fact that there is an extra cost imposed upon a .

When $H_{a:b} < 1$ the interaction is *relationship positive*. In this case, agent a needs to request background knowledge along with the principal message from b . The positivity is associated with the fact that there is an extra benefit for a , in that its knowledge will be expanded by interacting with b ¹.

Definition 3: For interacting agents a and b with respective internal models M_a and M_b , with respect to a given message m , the **relationship distance** between the agents $D_{a:b}$ is $f(M_a) - f(M_b)$, where $f(x)$ is some well-defined function on M_x .

According to this definition, identical agents would have zero distance. Dissimilar agents have a finite distance and one would also expect that the cost of communicating between them is a function of that distance i.e.:

When $D_{a:b} = 0$ the interaction is *relationship neutral*.

When $D_{a:b} > 0$ the interaction is *relationship negative*.

When $D_{a:b} < 0$ the interaction is *relationship positive*.

Homophily and relationship distance are clearly closely related. For example, one natural interpretation of the two is to see both as a statistical estimates of a model i.e. $f(x)$ is a probability distribution function describing M_x . Further, from basic information theory, the information contained in an event is defined as the log of its probability i.e. $I_x = \log p(x)$. We could then define:

$$\begin{aligned}
 H_{a:b} &= p(M_a) / p(M_b) \\
 \log H_{a:b} &= \log [p(M_a) / p(M_b)] \\
 \log H_{a:b} &= \log p(M_a) - \log p(M_b) \\
 &= D_{a:b} \\
 \log H_{a:b} &= D_{a:b}
 \end{aligned} \tag{1}$$

¹ Receiving knowledge is not always 'positive' in any absolute sense. There are costs associated with receiving, processing, and maintaining knowledge, as we shall see later.

Thus $D_{a:b}$ can be interpreted as an information distance, directly derivable from $H_{a:b}$, which is some statistical representation of the content or behaviour of a given model. This definition brings with it all of the nice properties that come with information measures. However, it is worth emphasising that there are many possible interpretations of $f(x)$, not just this probabilistic one, with consequences for how one interprets the relationship between $H_{a:b}$ and $D_{a:b}$.

While distance or homophily can give us an indication of the likely costs and degree of success of an interaction, it may be necessary to actually examine the specifics of the knowledge that agents share. In linguistics, the shared knowledge between agents is called the common ground (Clarke, Brennan, 1991; Clarke, 1996), and we can try to operationalise that notion:

Definition 4: For interacting agents a and b with respective internal models M_a and M_b , the **common ground** $G_{a:b}$ between them is the intersection of their model spaces $M_{a \cap b}$.

Note that we are saying nothing at present about the way an individual concept might be represented internally within an agent. For knowledge to appear in the common ground, both agents have to be able to recognise that the other agent possesses knowledge that is functionally equivalent to their own for a given task and context.

Once common ground is known, we can begin to examine its effect upon message transmission. An idealised message can be thought of as containing two parts. The first is a model, and the second is data encoded according to that model (Wallace and Boulton, 1968). Thus we can define a message from agent a to b as having the idealised form:

$$m_{a \rightarrow b} = M_{a \rightarrow b} + (D | M_a) \quad (2)$$

The model $M_{a \rightarrow b}$ is that subset of agent a 's overall knowledge that it has determined is needed to interpret the message data D . In information theory, this model would be the code used to encrypt data.

However, since some of the model $M_{a \rightarrow b}$ is known or is assumed to be shared between agents, not all of it needs to be transmitted. The optimal model to be transmitted t will be:

$$t(M_{a \rightarrow b}) = \{ \forall x: x \in M_{a \rightarrow b} \wedge x \notin M_{a \cap b} \} \quad (3)$$

and the actual message sent will be:

$$t(m_{a \rightarrow b}) = t(M_{a \rightarrow b}) + (D | M_a) \quad (4)$$

The effort required to transmit a particular message between two agents can be represented by the degree of message shortening that occurs because of the shared common ground between agents:

Definition 5: For interacting agents a and b with respective internal models M_a and M_b , the **grounding efficiency** for a message m sent from a to b is the ratio of the lengths of the actual message transmitted over the idealised message length determined by a :

$$\begin{aligned} G_E &= \frac{t(m_{a \rightarrow b})}{m_{a \rightarrow b}} \\ &= \frac{t(M_{a \rightarrow b}) + (D | M_a)}{M_{a \rightarrow b} + (D | M_a)} \end{aligned} \tag{5}$$

Grounding efficiency thus gives us some measure of the resource requirements for an interaction between two agents, all else being equal. If the average grounding efficiency is calculated by summing over a number of interactions, it gives the efficiency likelihood of a future interaction, based upon past experience i.e.:

$$G_{E(av)} = \frac{1}{n} \sum_n \frac{t(m_{a \rightarrow b})}{m_{a \rightarrow b}} \tag{6}$$

One of the reasons agents create common ground is to optimise their interaction. By sharing ground, less needs to be said in any given message, making the interaction less costly. Consequently, the goal in grounding is to develop sufficient ground to accommodate the anticipated efficiency requirements of future interactions. One can identify three different classes of interaction, based upon grounding efficiency:

- When $G_E < 1$, the interaction is *ground positive* since some shortening of the message has occurred.
- When $G_E = 1$ the interaction is *ground neutral*. The message in this case is simply sent in its idealised form.
- When $G_E > 1$, the interaction is *ground negative*. This means that the message sent is longer than the idealised length.

The effect of context upon interaction

The idea of the context within which an interaction occurs can be interpreted widely. Here it is understood to be the set of constraints upon communication imposed by the environment that is external to agents. Specifically, context imposes limits upon the physical resources available for communication. Examples of contextually constrained resources include time, bandwidth and money.

A channel forms part of the environment and limits the nature of any given communicative interaction. For example, when a channel mediates an interaction, we should be able to relate message-grounding requirements between agents to the transmission capacity of the channel that could successfully mediate their exchange:

Definition 6: For interacting agents a and b , a message m transmitted over a time t with a grounding efficiency of G_E , the **requisite channel capacity** C is:

$$C = \frac{m}{t} G_E$$

Recalling that average grounding efficiency is based upon the history of previous messages, making the channel capacity for the $n+1$ th message dependent upon the grounding history of the previous n messages:

$$C = \frac{m_{n+1}}{tn} \sum_{x=1}^n \frac{t(m_x)}{m_x} \quad (7)$$

With this equation, we connect together the triad of channel characteristics, context-specific message requirements and the nature of the relationship between communicating agents. It allows us to make three different types of inference:

- *Channel-centric inference*: For a given pair of agents and a message, we can specify channel capacity requirements.
- *Relationship-centric inference*: For a given agent, channel and time resource, we can say something about the common grounding that would help select an appropriate second agent to communicate with.
- *Task-centric inference*: Time can be considered a key task-limiting resource. For a pair of agents and a given channel, we can say something about the complexity of messages that can be exchanged between them over a given period of time.

The last of these suggests that we can generalise the previous equation further to incorporate other resources. Resource-specific versions of the equation can be crafted, substituting bandwidth for other bounded resources like time, monetary cost, utility of knowledge received etc. A general resource function, probably based upon utility, may be useful.

Grounding estimates lead to communication errors and inefficiencies

When agents create a message to suit a given context's needs, it must make guesses. For example, Equation 4 shows that a guess needs to be made about what model $t(m_{a \rightarrow b})$ is to be transmitted in a message. This guess is based upon estimates ϵ of what model is needed to explain the data $\epsilon(M_{a \rightarrow b})$, and what part of that model is already shared $\epsilon(M_{a \cap b})$. Thus grounding efficiency will always be a local measure when individual agents make it, since it is based upon local estimates of what needs to be transmitted:

$$t_a(M_{a \rightarrow b}) = \{ \forall x, x \in \epsilon(M_{a \rightarrow b}) \wedge x \notin \epsilon(M_{a \cap b}) \} \quad (8)$$

Consequently, errors occur in estimating both the common ground and the model needed for data interpretation. For example, an agent may end up explaining more of the model than is needed because it underestimates the common ground it shares with another agent, or conversely fails to explain enough. It may also simply lack the knowledge needed to interpret the data it sends. The possible error cases are summarised in Table 1.

	$< \epsilon_b (M_{a \cap b})$	$= \epsilon_b (M_{a \cap b})$	$> \epsilon_b (M_{a \cap b})$
$\epsilon_a (M_{a \cap b})$	<i>a confuses b:</i> <i>a overestimates common ground with b - model message too short and b cant fully interpret</i>	<i>a informs b:</i> <i>a's estimate of common ground agrees with b - model message optimal</i>	<i>a bores b:</i> <i>a underestimates common ground with b- model message too long, telling b what it already knows</i>

Table 1: Communication errors based upon common ground estimates.

With the notion of common ground, we are thus able to move away from the rather abstract idea of homophily, and begin to explore more precisely how relationship as modelled by shared knowledge affects communication. It also begins to give structure to the different types of interaction that can occur between agents, and the errors, misunderstandings and inefficiencies that result.

The Costs of Grounding

One consequence of agents having independent world models is the behaviour of *grounding* where agents devote some portion of their total interactions to checking if they understand each other (McCarthy & Monk, 1994). Grounding is the process in which agents specifically attempt to test the common language, knowledge and assumptions between them, and then attempt to modify deficiencies in their shared understanding to ensure the overall goal of an interaction succeeds.

We can think of grounding as a process for operationally testing and then ensuring M_1 and M_2 are sufficiently similar, within the context of a particular set of interactions, to guarantee a common interpretation of shared messages. Work in the psychological literature examines how people first establish, and then maintain, common ground whilst communicating (e.g. Clarke & Brennan, 1991).

Grounding is not a neutral process, but consumes some of the resources agents have available to communicate. These costs come in two varieties. Firstly, there are costs associated with interpreting the messages created in the process of grounding. Secondly, there are costs associated with maintaining the common ground itself, which we have characterised here as a common set of interpretive models of the world. These two costs have very different characteristics in relation to when they are incurred and how agents might choose to manage them, as we shall now see.

The Cost of Communicating Common Ground

If agents had identical models of the World, then the only costs of communication would arise out of transmitting across a noisy channel, and constructing and decoding messages according to shared protocols. However, when there is uncertainty about shared understanding, both sender and receiver incur extra costs during communication.

We can relate this grounding cost to the difference in the models between two communicating agents. For any given message, there may exist some additional background knowledge B that is possessed by the sending agent, and needed by the receiving agent to fully interpret its meaning. For example, if we tell someone to meet us at a certain location, but they do not know where it is, then we will also need to tell them additional information about its geographical situation and navigating a route there.

For a sending agent A_1 , costs are incurred in inferring B based upon its best understanding of the other agent's knowledge and the progress of the conversation, and then communicating B . There is clearly some monotonically increasing relationship between the cost of interaction for A_1 and the size of B . The more background knowledge we need to impart, the more effort it will require.

Equally A_2 bears reciprocal costs in receiving B in addition to the actual message. Again, there is an increasing set of costs associated with receiving, checking and testing B , as the size of B grows.

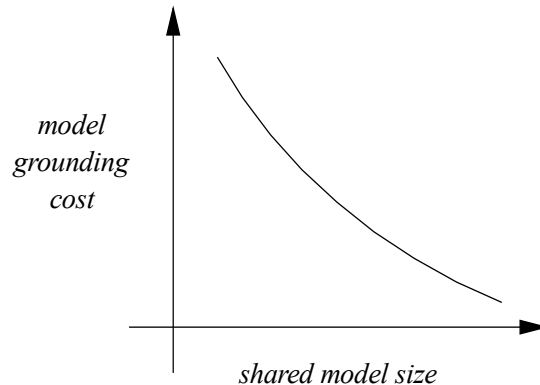


Figure 2: The cost of grounding models decreases with the size of the shared model space between two agents.

So, for any two agents participating in an interaction, the cost of grounding rises with the difference in their models of the world. Conversely, and more directly, the cost of communication decreases as the degree of model sharing increases (Figure 2).

The Cost of Maintaining Common Ground

Agents interact in a dynamic world, and their knowledge of the world needs to be refreshed as the world changes around them. Thus the accuracy of any model possessed by an agent decays over time. For a model to retain its accuracy, it must be maintained at a rate commensurate with the rate of world change. It follows that there is an on-going requirement in any dynamic universe, for agents to update shared models.

Keeping knowledge up-to-date requires effort on the part of an agent, and this effort is related to the size of the up-date task. If we think of these shared models as formally represented structures, then we know that the maintenance and updating costs are exponential in shape, based on the costs of writing, de-bugging and updating software (Littlewood, 1987). In fact, experience with software development shows that maintenance costs come not just from a need to keep up-to-date, but from the inherent flaws in the model building process. Our models are inherently flawed and ‘de-bugging’ costs rise with the size of a program. For the present argument, we only need note that many forces conspire to produce an increasing cost of model maintenance with model size. Consequently, the greater the size of the shared model of the world between two agents, the greater the need for updating and maintaining that model (Figure 3).

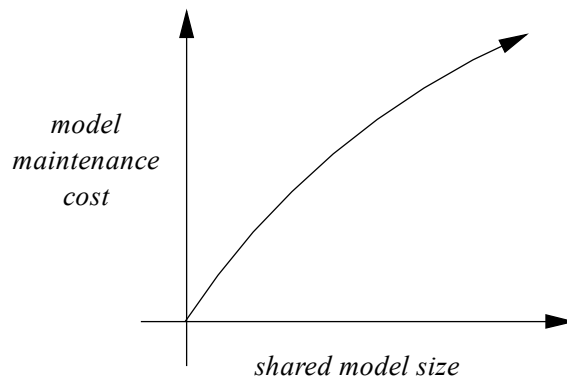


Figure 3: The cost of maintaining shared models increases with the size of the shared model space between agents.

Interactions are shaped to minimise grounding communication and maintenance costs

These two different costs associated with common ground give agents choices to either share mutual knowledge pre-emptively in anticipation of a communication, or just-in-time to satisfy the immediate needs of an interaction (Coiera, 2000). Since agents operate with finite resources, they will seek to minimise these costs on their resources. Firstly, agents have an incentive to minimise how much knowledge they share with each other, because maintaining and updating this shared body of knowledge is expensive. Intuitively we can think of the cost we incur 'keeping in touch' with people, and note that most people devote effort here commensurate with the value they place on the relationship and the amount of resource they have available for communication. In contrast, the less common ground agents share during a particular conversation, the more expensive that individual conversation will be, since additional knowledge will need to be communicated at the time. Intuitively, we can think of the effort one spends 'catching up' with an individual we haven't spoken with for some time, prior to getting to the point of a particular conversation. There are thus opposing incentives to minimise sharing of models prior to a conversation, and maximise the sharing during a conversation (Figure 4).

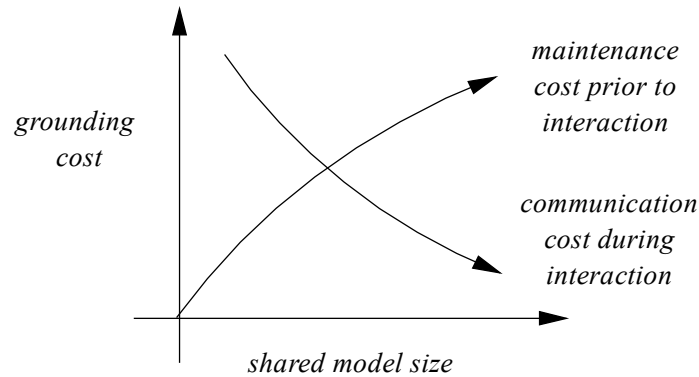


Figure 4: Grounding costs at the time of a conversation decrease with shared models, but carry an ongoing up-date burden that increases with the size of the shared space between agents.

The total cost of grounding T is therefore the sum of costs during an individual set of interactions D and the model maintenance efforts prior to them P i.e. $T = D + P$. Interestingly, this line of reasoning predicts that the total interaction costs have a minimum at an intermediate level of model sharing, where the cost curves cross (Figure 5). Given a desire to minimise costs by individual agents, then it should also be predictable that they will seek to find such intermediate levels of model sharing since they are the most economical over a long sequence of interactions.

Fundamentally, every agent makes some assessment of the costs $C(x)$ and benefits $B(x)$ of a potential interaction x , and when the benefits exceed costs, elects to interact (Frank, 1998) i.e.:

if $B(x) > C(x)$ then do x else don't

What emerges then is a picture of a series of cost-benefit trade-offs being made by individual agents across individual interactions, and populations of interactions. The goal is to maximise the utility of any given interaction, through a maximisation of benefit, and minimisation of cost. The result of these choices is that agents try to choose a degree of shared common ground with other agents that minimises interaction costs, but is still sufficient to accomplish the goal associated with interactions. This leads to a hypothesis:

Hypothesis 1: The *law of the mediated centre* states that an interaction system of communicating agents will be driven to an intermediate level of model sharing over a series of interactions where the total costs of maintaining shared models and communicating models at interaction time are minimised. The mediated centre represents an interaction equilibrium point.

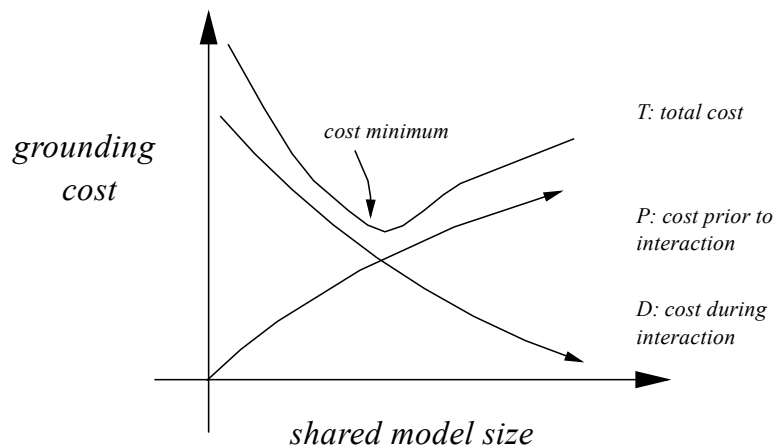


Figure 5: Total grounding costs are based on costs during and prior to a conversation, and reach a minimum at an intermediate level of model sharing between agents.

Many factors modify decisions to model-share, in particular an individual agent's assessment of the duration of its relationship with another agent, and the costs of being caught short during an individual conversation. Thus, if an interaction with another agent is infrequent, then it may be cheaper to choose to ground only during individual interactions. If the agent is going to have a series of similar interactions, then it may be cheaper in the long run to establish a high level of shared common ground, and minimise grounding during individual interactions. Equally, if individual interactions are resource critical, for example being time-bound, then there may be little opportunity to ground at the time of an interaction. Consequently, even if interactions are infrequent, the cost of failure to communicate may make it important to establish prior common ground.

Similarly, there are costs associated with the channel chosen for an interaction. Some channels may allow rapid and high-quality transmission of a message, but be very expensive. Others may be slow but cheap, and in many circumstances sufficient for the task at hand. Channel costs are numerous and include:

- the actual price charged for using the channel
- the effect of noise on the transmitted message
- the effort involved in locating channel access
- the number of connection attempts required before another agent replies
- the opportunity costs involved in message delay due to the channels transmission characteristics like bandwidth and latency
- and similarly, the time it takes to transmit a message

Conclusion

In this paper, a framework for defining interactions between agents with only bounded mutual knowledge has been developed, extending upon earlier work (Coiera, 2000). The development of the theory to incorporate sequences of interactions between individual

agents has lead to the notion of utility maximisation by agents, where they seek to optimise the degree of common ground between themselves.

The goal of developing such a mediated communication theory is to provide quantitative mechanisms for the design of interactions by individual computational agents, or for the design of interactions systems for use in complex organisations.

References

C R. Berger, Chatauqua: Why are there so few communication theories?, *Communication Monographs*, 58, 101-113, (1991).

H S. Bierman, L. Fernandez, *Game Theory with economic applications*, Addison-Wesley, Reading., Ma., (1995).

R Brooks, Intelligence without representation, *Artificial Intelligence*, 47, 139-159, (1991).

D Clarke, S. Brennan, Grounding in Communication, in L.B. Resnick J. Levine, & S.D. Behreno (Eds.), *Perspectives on socially shared cognition*, American Psychological Association, Washington, (1991).

H. Clark, *Using Language*, Cambridge University Press, Cambridge, (1996).

P. Cobley (ed.), *The Communication Theory Reader*, Routledge, London, (1996).

E. Coiera, When Communication is better than Computation. *Journal American Medical Informatics Association*. 7, 277-286, 2000.

R H. Frank, *Microeconomics and behavior*, 3rd Ed., McGraw Hill, New York, (1997).

R. Herken (ed.), *The Universal Turing Machine - A half-century survey*, Oxford University Press, Oxford, (1988).

P. F. Lazarsfeld, R. K. Merton, Friendship as Social process: a substantive and methodological analysis, in M. Berger et al. (eds.), *Freedom and Control in Modern Society*, New York, Octagon (1964).

B. Littlewood, *Software Reliability: achievement and assessment*, Blackwell Scientific Publications, (1987).

B. Moulin, B. Chaib-Draa, An overview of distributed artificial intelligence, in GMP O'Hare and NR Jennings (eds), *Foundations of Distributed Artificial Intelligence*, John Wiley and Sons, Inc., (1996).

J. C. McCarthy, A. F. Monk, Channels, conversation, co-operation and relevance: all you wanted to know about communication but were afraid to ask, *Collaborative Computing*, 1:35-60 (1994).

B. Nardi (ed), *Context in Consciousness - activity theory and human-computer interaction*, MIT Press, Cambridge Ma., (1996).

E. M. Rogers, *Diffusion of Innovations*, 4th Ed., Free Press, New York, (1995).

S. J. Russell, P. Norvig, *Artificial Intelligence - a modern approach*, Prentice Hall, Upper Saddle River, (1995).

L. Suchman, *Plans and situated actions - the problem of human-machine communication*, Cambridge University Press, Cambridge, (1987).

J. C. A. van der Lubbe, *Information Theory*, Cambridge University Press, (1997).

CS Wallace, DM Boulton, An information measure for classification, *Comp J*, 11, 185-195, (1968).