Reasoning About The Effects of Communication On Beliefs

R. Michael Young
Department of Computer Science
Stanford University
Stanford, CA 94305

May 26, 1994

Abstract

Perrault[2] has presented a formal framework describing communicative action and the change of mental state of agents participating in the performance of speech acts. This approach, using an axiomatization in default logic, suffers from several drawbacks dealing with the persistence of beliefs and ignorance over time. We provide an example which illustrates these drawbacks and then present a second approach which avoids these problems.

This second approach, an axiomatization of belief transfer in a nonmonotonic modal logic of belief and time, is a reformulation of Perrault's main ideas within a logic which uses an ignorance-based semantics to ensure that ignorance is maximized. We present an axiomatization of this logic and describe the associated techniques for nonmonotonic reasoning. We then show how this approach deals with inter-agent communications in a more intuitively appealing way.

1 Introduction

Speech Act Theory [4] focuses on modeling the change of mental state resulting from the performance of communicative action. This change in mental state typically involves the transfer of belief from one agent to another. For instance, if Karen tells Susan that it is raining outside, Susan should come to believe that Karen believes it is raining, and Susan herself may subsequently adopt the belief that it is raining. It is necessary that any formal approach to modeling this phenomenon be nonmonotonic, since the majority of the belief transfers which occur are themselves default inferences made by the hearer based on his or her beliefs regarding the speaker's beliefs.

The nonmonotonic approach to modeling communicative action was first attempted by Perrault [2] using Reiter's default logic [3]. A technical problem exists in the approach used by Perrault which results from the failure to include axioms affecting the default persistence of ignorance over time. Thus an agent's theory may support the fact that the agent is ignorant of some fact at time t, that is, $\neg B_{x,t}\varphi$, and yet be unable to deduce $\neg B_{x,t+1}\varphi$, that is, that the agent is still ignorant of the fact at time $t+1$ even though no new information about φ was received between times t and $t+1$.

The addition of a default axiom to allow persistence of ignorance in this framework causes further difficulty due to the interaction between it and default rules that refer to the belief state of the agents themselves. Perrault's default rule schemas, typically of the form

$$\frac{\alpha : MB_{x,t}\varphi}{B_{x,t}\varphi}$$

are defeated not only by $B_{x,t}\neg\varphi$, that is, belief in the negation of the proposition in question, but also unfortunately by *ignorance* of φ, or $\neg B_{x,t}\varphi$. In Section 3 we introduce a logic that specifically addresses the issues of ignorance and reasoning with incomplete or partial knowledge in a manner that provides for the persistence of ignorance and yet avoids this undesirable defeat property.

2 Perrault's Approach

Here we show how Perrault's approach deals with a communications scenario where our intuitions are relatively clear on what mental states should arise as a result of the communicative actions we describe. We highlight the dependence of this approach upon the lack of persistence of ignorance and then show how allowing default persistence creates new problems for this formalism. In Section 4.2 the same scenario will be cast in the framework of the axiomatization of communicative action in a logic of our own definition. We then show that the properties of this new axiomatization match our intuition for this class of inter-agent communication.

For completeness we provide a brief description of Perrault's axiomatization in the original default logic formalization. The reader should consult [2] for a full account of Perrault's work. Note that the formula $B_{x,t}\varphi$ is intended to mean that agent x believes formula φ at time t, the formula $I_{x,t}\varphi$ is intended to mean that agent x is ignorant of φ at time t[1], the formula $DO_{x,t}\alpha$ is intended to mean that agent x performs some action α at time t and the formula $Utter(\varphi)$ represents the act of uttering a sentence with propositional contents φ. Perrault's original axiom schema are as follows:

Consistency $$B_{x,t}\varphi \supset \neg B_{x,t}\neg\varphi \quad (1)$$

Closure $$B_{x,t}\varphi \wedge B_{x,t}(\varphi \supset \psi) \supset B_{x,t}\psi \quad (2)$$

Positive Introspection $$B_{x,t}\varphi \supset B_{x,t}B_{x,t}\varphi \quad (3)$$

Negative Introspection $$\neg B_{x,t}\varphi \supset B_{x,t}\neg B_{x,t}\varphi \quad (4)$$

Necessitation $$B_{x,t}\varphi, \text{where } \varphi \text{ is a tautology} \quad (5)$$

Memory $$B_{x,t}\varphi \supset B_{x,t+1}B_{x,t}\varphi \quad (6)$$

[1]Although Perrault did not define an ignorance operator, we provide the operator I here for clarity. An agent will be ignorant of some formula φ precisely when it neither believes φ nor it's negation. That is, $I_{x,t}\varphi \equiv \neg B_{x,t}\varphi \wedge \neg B_{x,t}\neg\varphi$.

Persistence $$B_{x,t+1}B_{x,t}\varphi \supset B_{x,t+1}\varphi \tag{7}$$

Observability $$DO_{x,t}Utter(\varphi) \wedge DO_{y,t}Obs(x) \supset B_{y,t+1}DO_{x,t}Utter(\varphi) \tag{8}$$

Belief Transfer Rule Schema $$\frac{B_{x,t}B_{y,t}\varphi : MB_{x,t}\varphi}{B_{x,t}\varphi} \tag{9}$$

Declarative Rule Schema $$\frac{DO_{x,t}Utter(\varphi) : MB_{x,t}\varphi}{B_{x,t}\varphi} \tag{10}$$

Perrault also has a default schema, intended to capture the closure of beliefs under defaults:

Default Closure Rule Schema

$$\text{If } \frac{\alpha : M\varphi}{\varphi} \text{ is a default rule then so is } \frac{B_{x,t}\alpha : MB_{x,t}\varphi}{B_{x,t}\varphi} \tag{11}$$

The specific example we wish to analyze involves two agents, S and H. Agent S will communicate to agent H that some sentence φ is true. Now it so happens that H has no *a priori* belief about φ or about S's beliefs about φ. Our intuition in this situation tells us that in the absence of any beliefs about φ, H would be safe in assuming that φ is true until some observation is made to the contrary.

Let the set A contain all instances of Perrault's axiom schemas. Then Perrault's approach starts with the following set of sentences:

$$W = A \cup \{I_{H,1}\varphi, I_{H,1}B_{S,1}\varphi, I_{H,1}\neg B_{S,1}\varphi, DO_{S,1}Utter(\varphi), DO_{H,1}Obs(S)\}$$

Perrault's Default Rule (10) and Default Schema (11) combine to give us the default

$$\frac{B_{H,2}DO_{S,1}Utter(\varphi) : MB_{H,2}B_{S,1}\varphi}{B_{H,2}B_{S,1}\varphi} \tag{12}$$

Application of this rule is the only way for $B_{H,2}B_{S,1}\varphi$ to enter any extension of W. An application of Perrault's Belief Transfer Rule then allows S to adopt belief in φ, as desired.

An unpleasant property of Perrault's axiomatization is the lack of persistence of ignorance. While Perrault's Persistence (7) and Memory (6) axioms ensure that a belief held at one time will continue to be held at all subsequent times, there are no analogous axioms to allow ignorance to persist. In point of fact we would not want ignorance to persist in every case, since that would prohibit our agents from gaining any new beliefs, but we do want ignorance to persist *as a default*. That is, an agent that is ignorant of a fact at a certain time should remain ignorant of that fact at all subsequent times unless there is some reason for it not to.

In this example the hearer is ignorant of φ at time point 1, that is $I_{H,1}\varphi \in W$. The only axioms that deal with transferring beliefs across time are the Persistence and Memory axioms, but these deal only with formulas of the form $B_{x,t}\varphi$, not those formulas expressing ignorance of the form $\neg B_{x,t}\varphi$, and so there is no way for $I_{H,2}\varphi$ to enter any extension of W.

If we were to attempt to solve this problem by including in A a default for the persistence of ignorance, multiple extensions of W would arise. In some extension E' of W, ignorance would not persist because of the default adoption of belief *before* the application of the ignorance persistence default. In some other extension E'', an instance of the default schema (12) above would be defeated due to the presence of $I_{H,2}B_{S,1}\varphi$ in E''. $I_{H,2}B_{S,1}\varphi$ would be in E'' just because H's ignorance *would* persist. This ignorance is sufficient to defeat the chain of default rule applications necessary for H to come to believe φ in E''.

In the following sections we define a logic that not only allows explicit reasoning about ignorance, but also incorporates the default persistence of ignorance in the semantics of the logic itself. We will show how this logic deals with the communications example in a more intuitive way.

3 Description of the Logic TI

Here we describe a nonmonotonic modal logic of belief and time called the logic of temporal ignorance, TI. This logic is a variation of the logic defined in [5], however we adopt a syntax similar to that used by Perrault in order to make the comparison of this use of TI to Perrault's approach clearer. For the formal definition of TI and a further discussion of its properties, including proofs of soundness and completeness, refer to [7].

Several different expressive levels are present in the logic TI. At the lowest level, there exists a set of basic tenseless propositions called *proposition-types* which are used to describe qualities about the domain of discourse of our theories. These proposition-types are associated with *time points*[2] where they may or may not be said to hold, resulting in *temporal assertions*. The highest level of the logic includes a modal operator for describing the state of belief about these temporal formulas (or in general, *any* formulas) for a particular agent at a particular point in time.

For instance, we may have as proposition-types for our use propositions like $BLOCK - RED$, whose truth indicates that a particular block on a table top was colored red. This proposition-type says nothing about the *time* at which the block was red. In order to refer to temporal concepts, say for instance that the block was red at time 4, we need some way to associate this proposition-type with that time point. In TI this is expressed as the temporal assertion

$$TRUE(4, BLOCK - RED)$$

In order to describe which agents *believe* which facts at which times we need yet another addition to our logic. This is accomplished in the following way: if agent R believes at time 8 that the block was red at time 4, we write the formula

$$B_{R,8}TRUE(4, BLOCK - RED)$$

As in Perrault's logic, we write $I_{x,t}\varphi$, that is, agent x is ignorant of the formula φ at time t, as shorthand for $\neg B_{x,t}\neg\varphi \wedge \neg B_{x,t}\varphi$.

3.1 Axiom System for Belief and Time Points

For the logic TI we will use a temporally-indexed version of the logic $K45$, also known as weak $S5$. Thus we have the following axioms for the modal operator $B_{x,t}$:

[2]For simplicity's sake we limit our discussion to a logic which is temporally point-based. The logic described in [7] associates proposition-types with *intervals*.

$$B_{x,t}\varphi, \text{ where } \varphi \text{ is a tautology} \tag{13}$$

$$B_{x,t}\varphi \wedge B_{x,t}(\varphi \supset \psi) \supset B_{x,t}\psi \tag{14}$$

$$B_{x,t}\varphi \supset B_{x,t}B_{x,t}\varphi \tag{15}$$

$$\neg B_{x,t}\varphi \supset B_{x,t}\neg B_{x,t}\varphi \tag{16}$$

We also have the rules of inference *modus ponens* and generalization.

3.2 Nonmonotonic Reasoning in This Framework

It has been shown[6] that many nonmonotonic logics are reducible to logical systems that utilize special semantics to select specific desirable models for their theories. In general, there will be many models for any particular set of sentences of TI. Not only do models of TI assign truth values to temporal assertions, but also to modal formulas of arbitrary depth. For example, suppose for some theory Δ_1 we have the set of proposition-types $\{p, q\}$, the single agent R and $\Delta_1 = \{B_{R,1}TRUE(4, p)\}$. Among the many models for Δ_1 are models in which $B_{R,1}TRUE(4, q)$ is true, models in which $B_{R,1}TRUE(4, \neg q)$ is true, and models where $I_{R,1}TRUE(4, q)$ is true.

By defining an ordering on models according to some *preference criteria* it is possible to select for use with our theories the subset of these many models which is maximal in this ordering. In particular we will design these preference criteria so that the maximally preferred models are just those which are maximally *ignorant*. That is, for any theory and any formula φ, if that theory does not imply $B_{x,t}\varphi$, our maximally preferred model will be one in which $I_{x,t}\varphi$ is true. Thus, in the above example using Δ_1, our preferred model would entail $I_{R,1}TRUE(4, q)$.

Simply preferring maximally ignorant models is not enough to guarantee a unique model for any theory of TI. We need some technique for selecting a single model when several *equi-ignorant* models exist. For example, what if some theory Δ_2 contains the single sentence $B_{R,1}\neg\varphi \vee B_{R,2}\psi$? Two equally maximally ignorant models exist for Δ_2, one model in which $B_{R,1}\neg\varphi$ is true and $B_{R,2}\psi$ is not, and another in which $B_{R,2}\psi$ is true and $B_{R,1}\neg\varphi$ is not. Our preference criteria as stated provides no way to select between the two. We will use a chronological ordering technique similar to the logic of chronological ignorance [5] where the preferred model is the one in which belief is established *as late as possible*[3]. This reflects our intuition that facts come to be believed as the result of some actions in the world or reasoning processes in the head of an agent; belief should not "spring to life" any sooner than necessary.

Looking again at Δ_2, we now have a technique to choose between the two models. Since in the first model φ is believed at a time before the point at which ψ is believed in the second, the first model is less preferred than the second. Therefore, even though in both models the same amount of information is believed, our agent remains ignorant longer in the second model, and thus it is maximally preferred.

Let's look at an informal example of nonmonotonic reasoning using this preference semantics. We will again consider the case of a single agent R and two proposition-types p and q. The theory Δ will consist initially of the sentences $\Delta = \{B_{R,1}TRUE(2,p), B_{R,1}TRUE(2,p) \wedge \neg B_{R,1}TRUE(2,\neg q) \supset B_{R,2}TRUE(2,q)\}$. In every model of Δ, $B_{R,1}TRUE(2,p)$ holds. Since our preference criterion is in effect chronologically minimizing belief, and since belief in $TRUE(2,q)$ is not necessitated by our theory, we have that in our preferred model $\neg B_{R,1}TRUE(2,\neg q)$ holds. Thus $B_{R,2}TRUE(2,q)$ is also entailed in our preferred model.

Now suppose that we add the sentence $B_{R,1}TRUE(2,\neg q)$ to Δ. Then $B_{R,1}TRUE(2,\neg q)$ is true in all models of Δ, and we can no longer conclude $B_{R,2}TRUE(2,q)$ as we could when our theory was ignorant of $TRUE(2,q)$ at time point 1.

4 Reasoning About Communicative Actions

When reasoning about communicative action, we are concerned with the process by which one agent comes to adopt the beliefs of another agent over time as they communicate. Here we describe a default system in which a hearer adopts the belief φ of a speaker if the speaker utters a sentence with

[3] Chronological ordering in itself is still not sufficient to guarantee unique models for our theories. Logics of chronological ignorance must also place syntactic restrictions upon the sentences of their theories. In this paper we restrict our discussion to those sentences which conform to these restrictions. For a full explanation, see [5].

content φ[4], the hearer was listening to the speaker, and φ is *consistent* with the beliefs of the hearer.

The process by which this transference of belief occurs is similar to that of Perrault's, using two main nonmonotonic rules. The first default rule captures belief adoption, that is, our agents assume that when one agent utters some sentence φ that agent in fact believes φ. The second, belief transfer, captures the assumption that because an agent believes some other agent believes φ, it is safe to come to believe φ itself.

4.1 Axiomatization of Communication and Belief Transfer

The axioms described here provide the essential definitions required to describe transfer of beliefs, and were chosen to parallel the axiomatization of speech acts in terms of default logic presented by Perrault. We differ here in the use of TI as our logic, and in the modification of some of Perrault's action description operators for simplicity's sake. Specifically, where Perrault used the operator $Obs_x(y)$ to indicate that agent x was observing the action of agent y at some time, we will follow [1] and use $DO_{x,t}(Obs(DO_{y,t}\alpha))$.

Memory
$$B_{x,t}\varphi \supset B_{x,t+1}B_{x,t}\varphi \tag{17}$$

Observability
$$DO_{x,t}(Utter(\varphi)) \wedge DO_{y,t}(Obs(DO_{x,t}(Utter(\varphi)))) \supset B_{y,t+1}DO_{x,t}(Utter(\varphi)) \tag{18}$$

Belief Transfer
$$B_{x,t}B_{y,t'}\varphi \wedge \neg B_{x,t}\neg\varphi \supset B_{x,t+1}\varphi \tag{19}$$

Declaration
$$B_{y,t}DO_{x,t}(Utter(\varphi)) \wedge \neg B_{y,t}\neg B_{x,t}\varphi \supset B_{y,t+1}B_{x,t}\varphi \tag{20}$$

Persistence
$$B_{x,t+1}B_{x,t}\varphi \supset B_{x,t+1}\varphi \tag{21}$$

[4] In this discussion we will assume that whenever a speaker utters a sentence, that utterance was performed with communicative intent.

4.2 TI's Solution to the Scenario

Now let us consider the problem from Section 2 expressed in TI. Let A contain the axioms defined in Section 3.1. Our initial theory will contain

$$\Delta = A \cup \{DO_{S,1}(Utter(\varphi)), DO_{H,1}(Obs(DO_{S,1}(Utter(\varphi))))\}$$

Δ contains no information about H's beliefs, reflecting H's ignorance about almost everything.

Now given Δ and the Observability Axiom (18), we have by our minimization scheme

$$B_{H,1}DO_{S,1}(Utter(\varphi)) \tag{22}$$

because of H's ignorance of φ, our preference semantics gives us

$$I_{H,1}\varphi \tag{23}$$

then by the Declaration Axiom (20) we have

$$B_{H,2}B_{S,1}\varphi \tag{24}$$

Again, by H's ignorance, minimization yields

$$I_{H,2}\varphi \tag{25}$$

and then by the Belief Transfer Axiom (19) we have that

$$B_{H,3}\varphi \tag{26}$$

The process of belief transfer and adoption is very straightforward in this approach. The hearer is initially ignorant of φ and so upon hearing the speaker utter φ the hearer comes to believe that speaker believes φ (24). This belief and the hearer's ignorance of φ then allows the hearer to adopt φ as well (26).

5 Conclusions

We have shown that the particular default logic approach described in Sections 1 and 2 relies on a counter-intuitive property of non-persistence of ignorance, and cannot be corrected by a straightforward addition of default axioms. In the same case, ignorance of belief plays a critical role in the desired *adoption* of beliefs for the semantic minimization approach.

One possible criticism to this comparison would be to note that in the semantic minimization approach we established ignorance of a certain fact by failing to include formulas regarding that fact in an agent's theory and then relying on the semantics of the logic to select models ignorant of that fact. The same technique could be used for the default logic approach by prohibiting sentences describing the ignorance of certain facts from W. In this manner no default rules would be defeated by ignorance since ignorance would be not be provable. This approach, however, would mandate that ignorance *never* be provable, i.e., we would have to restrict ourselves to theories where ignorance is not logically entailed, thus no explicit reference to ignorance could be made. For instance, this would prohibit one agent communicating its own ignorance about some fact to another agent. In addition, any inference which was based on ignorance would have to be a default inference and thus one that was defeasible.

Unfortunately, ignorance is the norm rather than the exception when reasoning about the world. Declarative communication will typically occur when one agent wishes to inform another agent of some fact which the second agent is ignorant of. The inability of agents to reason *explicitly* about ignorance will cause considerable problems for systems using the default axiomatization which Perrault describes.

It is clear, however, that using a logic which makes explicit the notions of ignorance in its nonmonotonic behavior has no difficulty dealing with the loss of ignorance as a result of observing declarative communication.

6 Acknowledgements

I would like to thank Ray Perrault for providing the initial impetus for this work as well as supervising its development and participating in many useful discussions about the paper's contents. I would also like to thank Yoav Shoham, Fanghzen Lin, Susan Altman, Doug Appelt, Ken Fertig, Matt Ginsberg, Karen Myers, Nils Nilsson and Eunok Paek.

References

[1] Appelt, D. and Konolige, K., "A Practical Nonmonotonic Theory for Reasoning about Speech Acts," in *Proceedings of the 1988 Conference of the Association for Computational Linguistics.*

[2] Perrault, C. R., "An Application of Default Logic to Speech Act Theory," in *Intentions in Communication*, Cohen, R., Morgan, J. and Pollack, M., eds., MIT Press, Cambridge, MA, 1990.

[3] Reiter, R., *A Logic for Default Reasoning,* Artificial Intelligence, 1983.

[4] Searle, J. R., *Speech Acts: An essay in the philosophy of language*, Cambridge: Cambridge University Press, 1969.

[5] Shoham, Y., *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence,* Ph.D Thesis, Yale University, 1986.

[6] Shoham, Y. (87), *Nonmonotonic Logics: Meaning and Utility*, Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Milan, Italy, 1987.

[7] Lin, F., Shoham, Y. and Young, R.M., "Evolving Knowledge of Temporal Information", Stanford Computer Science Department Technical Report, in preparation.

[8] Young, R.M., "Reasoning About the Effects of Communicative Action on Belief," Stanford Computer Science Department Technical Report, in preparation.