On Clark and Schaefer's *Contribution* Model and its applicability to Human-Computer Collaboration

David R. Traum
UMIACS
University of Maryland
A. V. Williams Building
College Park, MD. USA
traum@cs.umd.edu

Abstract

I discuss the suitability of Clark and Schaefer's Contribution model of Grounding, as a design model for computer collaboration. This model is very appealing and, at a coarse level very influential and important for viewing dialogue as a collaborative process. While the main ideas are directly applicable and important considerations both for designers of collaborative systems and for those systems, themselves to attend to, the specific details are not particularly well suited to direct implementation.

1 Overview

The "Clark model" that I wish to discuss is Clark and Schaefer's Contribution model of Grounding, as laid out in [Clark and Schaefer, 1989]. This model is very appealing and, at a coarse level very influential and important for viewing dialogue as a collaborative process. While the main ideas are directly applicable and important considerations both for designers of collaborative systems and for those systems, themselves to attend to, the specific details are not particularly well suited to direct implementation. The main deficiencies are that this is not an on-line model of the grounding process, and offers little in the way of normative guidelines for what to do, or predictive power as to what to expect next in a collaborative dialogue.

The concept of *common ground*, or similar notions, such as *mutual belief* or *em a shared conception* have been very important in Cognitive Science theories of collaboration/cooperation.[1] Indeed, the two currently dominant accounts of intentional collaboration in AI, *Joint Intentions/Teamwork* [Cohen and Levesque, 1991b] and *SharedPlans* [Grosz and Sidner, 1990, Grosz and Kraus, 1993] both include *mutual beliefs* as key components in their definitions. However, few have taken helpful stands on *how* mutual belief is established in dialogue. Most agree that acknowledgment plays some role, but there are proofs that no amount of acknowledgments can insure perfect mutual belief in a noisy environment [Halpern and Moses, 1990]. Most AI dialogue researchers, including those above[2], have settled for the opposite extreme — that in virtue of copresence and sometimes other assumptions, information will automatically become part of common ground.[3]

[1] several authors have taken stands on distinguishing these concepts, e.g. [Roschelle and Teasley, 1995], however others use them interchangeable, or with opposite definitions *c.f.,* [Allwood, 1976]. I will not distinguish them here but will use the term collaboration without trying to engage in this ongoing debate.

[2] [Cohen and Levesque, 1991a] talked about the desire for mutual belief as a motivation to perform acknowledgments, though they did not provide a model of the role the acknowledgment was supposed to play in actually achieving this mutual belief

[3] Interestingly, this is based on a previous "Clark model", [Clark and Marshall, 1981].

The use of acknowledgments has also been seen in HCI systems, as well as human conversation, in a variety of ways. However there has been little systematicity in most systems as to when an acknowledgment should appear in a dialogue, leading to system behaviors which can be very frustrating to users (e.g., on a PC, being forced to click "okay" to an incomprehensible message before being allowed to continue.

Clark and Schaefer's model is one of the first and clearest of how this achievement of mutual understanding is itself a collaborative process, brought about by active participation by multiple participants. I include a brief summary of this model in the Appendix.

2 Deficiencies of the Contribution Model

Although the contribution model is perhaps the first explicit model of how grounding takes place and why acknowledgments occur, it still is lacking in a number of particulars, especially when used to design a collaborating system.

2.1 How much acceptance is enough?

Since Clark and Schaefer assert that each signal (including acceptances) must itself be a presentation which needs acceptance, it is not clear that contributions are *ever* really complete. For example, in the simplest case, *contributions by turns*, Speaker A's first turn is a presentation part of a first contribution. Speaker B's following turn is an acceptance part of that contribution, but also is the presentation part of a next contribution. What is unclear is whether this second utterance must be accepted in order to fulfill its acceptance function. If so, as Clark and Schaefer seem to imply, then not even the next utterance by A will completely ground the first contribution: this acceptance of the acceptance will itself need to be accepted, and so on, ad infinitum. If it is possible, as they suggest, that some acceptances need not be accepted themselves, then this opens the possibility that the acceptance part of an utterance need not be itself accepted (though any actual intended next contribution which is part of the same utterance would still have to be accepted). Some utterances which are merely acceptances might not be presentations as well. Once one accepts this basic premise, the question arises as to which kind of utterances do not need to be accepted. My own thinking on this subject has changed somewhat: originally, [Traum and Hinkelman, 1992], I followed the hypothesis that acknowledgments themselves did not need acknowledging, but that all "classical" speech acts did. Later, [Traum, 1994], I expanded this to include also other kinds of simple "acceptance" that operated on a level of agreement as well as that of understanding. A current hypothesis is that no *backward function* (i.e., those referring back to a previous act requires acknowledgment (however, the same utterance will often also have a forward function, which *will need to be acknowledged*. Empirical studies also show that not only the style, but also the frequency of acknowledgments vary widely, depending on the media involved [Traum and Heeman, 1997, Dillenbourg *et al.*, 1997].

A firm supporter of Clark and Schaefer's model might retort, that, all acceptances *do* require acceptance, but the strength of evidence principle takes care of this by allowing for diminishing degrees of evidence. A problem with this is that eventually it will have to bottom out to zero, and how does one distinguish "little or no evidence needed" from "evidence not needed"? This leads into my next criticism:

2.2 Problems with Graded evidence of Understanding

Clark and Schaefer's graded types of evidence, summarized in Table 1 has been adapted by some in the community. However it has several problems. First, there's not great evidence for the ordering itself. E.g., "demonstration" is actually usually greater evidence of understanding than "display", since it shows some level of understanding as well as just perception. A bigger problem, however, is with "initiation of next relevant contribution." The issue is when this can actually be seen as acceptance behavior, and when as

mere obliviousness - e.g., if the next contribution is just as relevant as if the putative contribution had never been made. It is often this case with acknowledgments, especially in the form of backchannels. Positive feed back in the form of backchannels is generally not itself acknowledged explicitly, so how does one tell whether continuing on is acceptance of the backchannel or not?

2.3 Off-line nature of phases

It is often hard to tell whether a particular utterance is part of the presentation phase or the acceptance phase. Self-initiated self-repair and other-agent completions are considered part of the presentation phase, but other-repair and other-initiated self-repair are part of the acceptance phase. Either one can have embedded contributions, in the form of insertion sequences or clarification sub-dialogues, so, in the case of an other-initiated self-repair, it's hard to tell whether it is part of the presentation phase or the acceptance phase. We often need to look at large segments of the conversation, both before and afterwards before deciding how a particular utterance fits in.

The model is thus of little use to an agent in the midst of a conversation deciding what to do next based on what has happened before. Realizing that a presentation has been made but has not yet been accepted can lead one to initiate the acceptance phase, but it's not clear when a presentation or acceptance is complete, or whether the knowledge of being in the presentation phase or acceptance phase has any consequences for what should be uttered.

From a processing point of view, the main deficiency of Clark and Schaefer's **contribution** model of grounding is that there is no easy way to tell the "state" of the current contribution while engaged in a conversation. Although we might represent a contribution as a transition network such as that in Figure 1, with a grounded contribution being one in the final state, F, this is not sufficient to monitor on-line conversation.

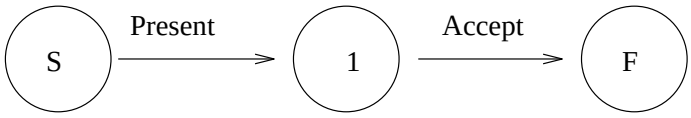

Figure 1: Transition Network for Contributions

We know that to start a contribution, a **presentation** must be performed primarily by one agent, whom we will call the *Initiator* (I) and then an **acceptance** must be performed primarily by the other agent, whom we will call the *Responder* (R), but what is less obvious is how to tell when a presentation has been performed. Another way of looking at this question is: given an utterance by the initiator, how does it function as part of the current contribution? Does it start, continue, or complete a presentation? Unfortunately, there is no way to recognize whether a presentation is complete, just by looking at an utterance itself. Consider the following sequences in examples (1) and (2).

	(1)	**(2)**
1	I: Move the boxcar to Corning	I: Move the boxcar to Corning
2		R: ok
3	I: and load it with oranges	I: and load it with oranges
4	R: ok	R: ok

Since, according to Clark and Schaefer, the sequence in example (1) is a single contribution, that means that the presentation phase must encompass both of the utterances by the initiator. However, in example (2),

there are two contributions with two separate presentations by I[4], and thus the first utterance by the initiator is a complete presentation. Since these sequences are identical up to the second utterance, there is, in general, no way to tell whether a presentation is complete until another action starts. This becomes a more serious matter because a contribution must be composed of actions by both participants, and thus there must be some way for the individual participants to determine an appropriate next action, given the current state.

3 Other models

While Clark and Schaefer's model is not ideally suited to the design of collaborative systems, I believe those that are better suited are those that are directly inspired by it (e.g., my own work) or very similar in spirit (e.g., [Allwood *et al.*, 1992]. In [Traum, 1994], I presented an on-line version of the contribution model, which was directly used as part of a collaborative dialogue agent. This model collapsed the different types of acceptance, but extended the building blocks of the units of common ground to those that could be realized with a single utterance, thus allowing an agent to track progress with each communication and without requiring lookahead.

Models of collaboration, such as those of Cohen and Levesque, or Grosz and Sidner will be necessarily incomplete if they use notions related to common ground without a good account of how it can be achieved. Collaborative agents which are built on these guidelines may run into trouble whenever they must communicate and communication is uncertain. Augmentations using ideas from Clark and Schaefer's contribution model can provide more complete models of agent collaboration.

4 Assessment

As computer collaborators become increasingly sophisticated and involved in important interactions with users, they will increasingly need to be able to reason explicitly about the state of common ground and whether this needs to be changed, and how. Simplistic techniques such as relying on the user to understand, or attempting to force understanding through "dialogue box clicking" will not be sufficient in all cases. However, reasoning about the principles involved in Clark's work, such as the grounding criterion, degree of grounding, and methods of increasing the degree of grounding will allow more natural and fluent interactions. Therefore, in spite of it's limitations as a "design model" (a task for which it was never intended), it can be very useful in helping designers develop suitable design models.

References

[Allwood *et al.*, 1992] Jens Allwood, Joakim Nivre, and Elisabrth Ahlsen. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9, 1992.

[Allwood, 1976] Jens Allwood. *Linguistic Communication as Action and Cooperation*. PhD thesis, Göteborg University, Department of Linguistics, 1976.

[Clark and Marshall, 1981] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, 1981. Also appears as Chapter 1 in [Clark, 1992].

[Clark and Schaefer, 1989] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989. Also appears as Chapter 5 in [Clark, 1992].

[4] Clark & Schaefer would also analyze the acknowledgments by R as the presentation phases of separate contributions.

[Clark, 1992] Herbert H. Clark. *Arenas of Language Use*. University of Chicago Press, 1992.

[Cohen and Levesque, 1991a] Phillip R. Cohen and Hector J. Levesque. Confirmations and joint action. In *Proceedings IJCAI-91*, pages 951–957, 1991.

[Cohen and Levesque, 1991b] Phillip R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 35, 1991.

[Dillenbourg *et al.*, 1997] P. Dillenbourg, P. Jermann, D. Schneider, D. Traum, and C. Buiu. The design of moo agents: Implications from a study on multi-modal collaborative problem solving. In *Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED 97)*, pages 15–22, 1997.

[Grosz and Kraus, 1993] Barbara [J.] Grosz and Sarit Kraus. Collaborative plans for group activities. In *Proceedings IJCAI-93*, pages 367–373, 1993.

[Grosz and Sidner, 1990] Barbara J. Grosz and Candace L. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[Halpern and Moses, 1990] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.

[Roschelle and Teasley, 1995] J. Roschelle and S.D. Teasley. The construction of shared knowledge in collaborative problem solving. In C.E. O'Malley, editor, *Computer-Supported Collaborative Learning*. Springer-Verlag, 1995.

[Traum and Heeman, 1997] David R. Traum and Peter Heeman. Utterance units in spoken dialogue. In Elisabeth Maier, Marion Mast, and Susann Luperfoy, editors, *Dialogue Processing in Spoken Language Systems — ECAI-96 Workshop*, Lecture Notes in Artificial Intelligence, pages 125–140. Springer-Verlag, Heidelberg, 1997.

[Traum and Hinkelman, 1992] David R. Traum and Elizabeth A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599, 1992. Special Issue on Non-literal language.

[Traum, 1994] David R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester, 1994. Also available as TR 545, Department of Computer Science, University of Rochester.

Appendix: Summary of Clark and Schaefer's model

It involves the augmentation of common ground as the production of *contributions*, composed of two parts. First, the contributor specifies the content of his contribution and the partners try to register that content. Second, the contributor and partners try to reach the *grounding criterion*, which Clark and Schaefer state as follows, "The contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose" [Clark and Schaefer, 1989, p. 262]. Clark and Schaefer divide the contribution into two phases as follows (for two participants, **A** and **B**) [Clark and Schaefer, 1989, p. 265]:

Presentation Phase: **A** presents utterance **u** for **B** to consider. He does so on the assumption that, if **B** gives evidence **e** or stronger, he can believe that **B** understands what **A** means by **u**.

Acceptance Phase: **B** accepts utterance **u** by giving evidence $\mathbf{e}'$ that he believes he understands what **A** means by **u**. He does so on the assumption that, once **A** registers evidence $\mathbf{e}'$, he will also believe that **B** understands.

Clark and Schaefer claim that once both phases have been completed, it will be common ground between **A** and **B** that **B** understands what **A** meant. Each element of the contribution may take multiple conversational turns and may include whole embedded contributions. Rather than a straightforward acceptance, **B** can instead pursue a repair of **A**'s presentation, or ignore it altogether. **B**'s next turn, whether it be an acceptance, or some other kind of utterance, is itself the presentation phase of another contribution. Thus **A** must accept **B**'s acceptance, and so on.

There are different types of evidence which can be given to show understanding. The main types considered by Clark and Schaefer are shown in Table 1, in order from strongest to weakest.

1	**Display**	B displays verbatim all or part of A's presentation.
2	**Demonstration**	B demonstrates all or part of what he has understood A to mean.
3	**Acknowledgment**	B nods or says "uh huh", "yeah", or the like.
4	**Initiation of relevant next contribution**	B starts in on the next contribution that would be relevant at a level as high as the current one.
5	**Continued attention**	B shows that he is continuing to attend and therefore remains satisfied with A's presentation.

Table 1: [Clark and Schaefer, 1989, p. 267]: Types of Evidence of Understanding

The strength of evidence needed for grounding depends on several factors, including the complexity of the presentation, how important recognition is, and how close the interpretation has to be. They try to avoid infinite recursion in accepting acceptances by invoking the following **Strength of Evidence Principle:** The participants expect that, if evidence $\mathbf{e_0}$is needed for accepting presentation $\mathbf{u_0}$, and $\mathbf{e_1}$ for accepting presentation of $\mathbf{e_0}$, then $\mathbf{e_1}$ will be weaker than $\mathbf{e_0}$.