Extended Abstract: Computational Models of Non-cooperative dialogue

David R Traum Institute for Creative Technologies
University of Southern California
Marina Del Rey, CA 90292 USA
traum@ict.usc.edu

Cooperativity is usually seen as a central concept in the pragmatics of dialogue. There are a number of accounts of dialogue performance and interpretation that require some notion of cooperation or collaboration as part of the explanatory mechanism of communication. For instance, Grice's cooperativity principle and associated maxims are used to explain conversational implicature (Grice, 1975). Searle uses general principles of cooperative conversation to account for indirect speech acts (Searle, 1975). Clark and Wilkes-Gibbs use a principle of "least collaborative effort" as a goal of the processes of grounding and accepting referring expressions. (Clark and Wilkes-Gibbs, 1986).

Alwood (Allwood, 1976) considers that full-blown communication requires at least some degree of cooperation, and defines ideal cooperation between a number of interacting normal rational agents as adherence to the following principles:

1. they are voluntarily striving to achieve the same purposes,
2. they are ethically and cognitively considering each other in trying to achieve these purposes.
3. they trust each other to act according to 1 and 2 unless they give each other explicit notice that they are not'. consisting of four parts:

Most advanced computational work on dialogue agents has also generally assumed cooperativity. Simple dialogue systems, e.g (Sutton et al., 1996), are programmed to react directly to specific types of inputs, without doing much pragmatic reasoning. Some advanced systems are formulated as agents that reason about attitudes such as belief, desire, and intention, e. g. (Cohen and Perrault, 1979; Allen and Perrault, 1980). These systems do means-ends reasoning to develop plans that further their goals which can be adopted as intentions, and also recognize the plans of others. There is still a tension between the model of individual agency and coordinated action, which is often modelled using principles of cooperativity, collaboration including such notions as joint intentions (Cohen and Levesque, 1991) and Shared Plans (Grosz and Sidner, 1990). These notions are used to automate the kinds of pragmatic reasoning described by Grice and Searle and compute speaker meaning using contextual knowledge as well as compositional semantics. This notion of cooperativity in conjunction with rational agency can be a powerful mechanism for allowing systems to engage in human-like flexible dialogues.

The cooperative principles are reasonable for the vast amount of domains that people have built dialogue systems for: service or information providing systems, in which the goals of both the system and user can be seen to coincide. What happens, though, when there is no shared goal, or cooperation breaks down in other ways, e.g., lack of cognitive or ethical consideration (and/or follow-through) or lack of trust? Many models have little to say about this kind of dialogue, and in fact disparage non-cooperative behavior in human-machine dialogue because it "easily leads to miscommunication and an unnecessarily long, complicated, and perhaps failed dialogue because of the system's limited abilities to detect, handle, and recover from a

Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue, June 2–4, 2008, London, U.K.

non-cooperative dialogue flow." (Klein et al., 1999; Hajdinjak and Mihelic, 2004). While cases of dialogue systems that are intentionally non-cooperative are not yet common for human-computer dialogue, there are a number of applications in which they are important, including:

- intelligent tutoring systems, e.g. (Zinn et al., 2002), in which the tutor must sometimes override the local desires of the student for the assumed greater good of education
- commercial bargaining agents, e.g., (Jameson et al., 1994; Jameson and Weis, 1995), in which the buyer and seller have opposite goals, at least in terms of price.
- more generally, assistant agents in which the agent may talk to someone other than its owner, in which it should take up the goals of its owner rather than others whom it may engage in dialogue with.
- role-playing training agents, in which the agents are playing roles which are not cooperative in order to let a user practice and learn how best to deal with such situations. (Traum et al., 2008; Traum et al., 2007)

In (Traum and Allen, 1994), we presented a model of coordinated dialogue behavior that did not rely on cooperativity for basic interaction. In that view, dialogue behavior could be motivated either by individual goals (which might or might not be shared or cooperatively adopted) or obligations, which are imposed by norms of social interaction and can be ignored only with potential social penalties. This had the potential to handle question-answering in non-cooperative situations, but the Trains system which used it (Allen et al., 1995) was highly cooperative.

More recently, we have been working on a number of virtual humans, who engage in face to face spoken dialogue and act as role-players for domains such as non-team negotiation (Traum et al., 2005), as shown in Figure 1, multiparty negotiation, as shown in Figure 2, and questioning interviews (Traum et al., 2007), as shown in Figure 3. In these domains, cooperativity is an achievement rather than an assumption. The agents can choose to be cooperative or uncooperative. Dialogue must proceed in

Figure 1: SASO-ST Negotiation in the Clinic: Dr Perez

both of these cases, and in fact, dialogue is one of the principal means of increasing cooperativity. We thus need accounts of aspects of dialogue behavior in which cooperativity does not play an essential role, as well as other computational mechanisms for specific uncooperative behaviors.

Figure 2: SASO-EN Negotiation in the Cafe: Dr Perez (left) looking at Elder al-Hassan

This talk will outline some cases of noncooperative communication behavior and computational dialogue mechanisms that can support these kinds of behavior, including generating, understanding, and deciding on strategies of when to engage in uncooperative behavios. Behaviors of interest include

- unilateral topic shifts or topic maintenance
- avoidance
- competition
- unhelpful criticism

- withholding of information or services
- lying & deception
- competition
- antagonism
- rejection of empathy

Figure 3: Tactical Questioning: Hassan

The decision of whether to be cooperative or not and how to behave in each case depends on a number of factors, including the standard notions of belief, desire, intention, obligation, and initiative, but also factors such as trust, solidarity, power, status, and respect.

We will present preliminary computational models of these factors and illustrate their use with examples of interactions with the characters shown in Figures 1, 2, and 3.

Acknowledgments

We would like to thank the rest of the Virtual Human team at USC. This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

James F. Allen and C. Raymond Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178.

James. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, and D. R. Traum. 1995. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7:7–48.

Jens Allwood. 1976. *Linguistic Communication as Action and Cooperation*. Ph.D. thesis, Göteborg University, Department of Linguistics.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39. Also appears as Chapter 4 in (Clark, 1992).

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

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

Phillip R. Cohen and C. R. Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

J. Paul Grice. 1975. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 41–58. Academic Press.

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

Melita Hajdinjak and France Mihelic. 2004. Information-providing dialogue management. In Petr Sojka, Ivan Kopecek, and Karel Pala, editors, *TSD*, volume 3206 of *Lecture Notes in Computer Science*, pages 595–602. Springer.

Anthony Jameson and Thomas Weis. 1995. How to juggle discourse obligations. In *Proceedings of the symposium on Conceptual and Semantic Knowledge in Language Generation*, Nov.

Anthony Jameson, Bernhard Kipper, Alassane Ndiaye, Ralph Schäfer, Joep Simons, Thomas Weis, and Detlev Zimmermann. 1994. Cooperating to be Noncooperative: The Dialog System PRACMA. In B. Nebel and L. Dreschler-Fischer, editors, *Proceedings of the Eighteenth German Conference on Artificial Intelligence (KI-94)*, pages 106–117. Springer.

Marion Klein, Niels Ole Bernsen, Sarah Davies, Laila Dybkjaer, Juanma Garrido, Henrik Kasch, Andreas Mengel, Vito Pirrelli, Massimo Poesio, Silvia Quazza, and Claudia Soria. 1999. Supported coding schemes.

Deliverable D1.1, MATE Project. available from http://www.dfki.de/mate/d11/.

John R. Searle. 1975. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics*, volume 3: Speech Acts, pages 59–82. Academic Press.

S. Sutton, D. G. Novick, R. A. Cole, and M. Fanty. 1996. Building 10,000 spoken-dialogue systems. In Proceedings 4th International Conference on Spoken Language Processing (ICSLP-96).

David R. Traum and James F. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 1–8.

David Traum, William Swartout, Stacy Marsella, and Jonathan Gratch. 2005. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *In proceedings of the Intelligent Virtual Agents Conference (IVA)*, pages 52–64. Springer-Verlag Lecture Notes in Computer Science, September.

David Traum, Antonio Roque, Anton Leuski, Panayiotis Georgiou, Jillian Gerten, Bilyana Martinovski, Shrikanth Narayanan, Susan Robinson, and Ashish Vaswani. 2007. Hassan: A virtual human for tactical questioning. In *The 8th SIGdial Workshop on Discourse and Dialogue*.

David Traum, William Swartout, Jonathan Gratch, and Stacy Marsella. 2008. A virtual human dialogue model for non-team interaction. In Laila Dybkjaer and Wolfgang Minker, editors, *Recent Trends in Discourse and Dialogue*. Springer.

Claus Zinn, Johanna D. Moore, and Mark G. Core. 2002. A 3-tier planning architecture for managing tutorial dialogue. In Stefano A. Cerri, Guy Gouardères, and Fábio Paraguaçu, editors, *Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 574–584. Springer.