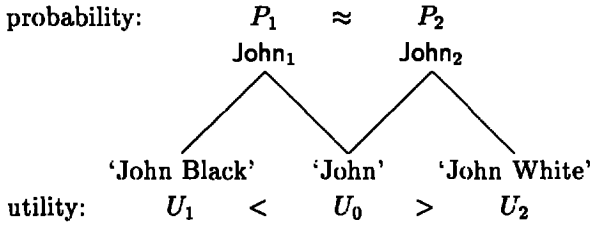
Optimality in Communication Games

HASIDA, Kôiti
Electrotechnical Laboratory
1-1-4, Umezono, Tukuba 305 Japan
hasida@etl.go.jp

Jerry R. Hobbs **Megumi Kameyama**
SRI International
333 Ravenswood Avenue, Menlo Park, CA 94025 U.S.A.
{hobbs, megumi}@ai.sri.com

A *meaning game* addresses a *turn* of communication $\langle c_S, m, c_R \rangle$, which stands for a course of events where a sender S, intending to communicate semantic content c_S, sends a message m to a receiver R and R interprets m as meaning c_R. Hasida (1996), extending Hasida, Nagao, & Miyata (1995), claims that natural-language meaning games are played at their Pareto-optimal equilibria. Here we revise this claim. See Hasida (1996) for related work, terminology, literature, etc.

Consider the meaning game depicted below. Here S



wants to refer to either John_1 or John_2, John_1 may be called either 'John Black' or just 'John,' and John_2 may be called either 'John White' or just 'John.' As indicated in the figure, let us assume that the prior probabilities P_1 and P_2 of references to John_1 and John_2 are nearly equal. 'John' incurs clearly smaller cost of utterance and interpretation than 'John Black' and 'John White.' So $U_1 < U_0 > U_2$, where U_i is the utility (negative cost) of the corresponding expression. Let us further assume $U_1 \approx U_2$.

There are three equilibria which guarantee successful communication:

(1) S means John_1 by saying 'John Black' and John_2 by saying 'John.' R interprets 'John' as meaning John_2.

(2) S means John_1 by saying 'John' and John_2 by saying 'John White.' R interprets 'John' as meaning John_1.

(3) S says 'John Black' to mean John_1 and 'John White' to mean John_2. We do not care how R might interpret 'John.'

In all the cases, R interprets 'John Black' as meaning John_1 and 'John White' as meaning John_2.

The expected utilities (apart from success of communication) associated with these equilibria are $E[1] = P_1U_1 + P_2U_0$, $E[2] = P_1U_0 + P_2U_2$, and $E[3] = P_1U_1 + P_2U_2$. Note $E[1] > E[3]$ and $E[2] > E[3]$. Namely, (1) and (2) are Pareto superior to (3). However, it is (3) that people tend to settle on in this meaning game.

This is because common knowledge is lacking by which S and R can choose the same equilibrium from (1) and (2). If they had common knowledge about enough detail of the whole game (P_i and U_i) and one of the two equilibria were Pareto superior to the other (hence being the unique Pareto-optimal equilibrium), then the players would be able to commonly adopt that equilibrium, because in that case they would commonly know it to maximize their expected utilities.

Let us conjecture the following.

(4) The solution (if any) of a natural-language meaning game is the Pareto optimum among the equilibria which are commonly Pareto comparable with every other equilibrium.

We say two equilibria are commonly Pareto comparable when it is commonly known that one of them is Pareto superior to the other. Note that, in the current example, only (3) is commonly Pareto comparable with the other equilibria.

In order to deal with such a case, metareasoning about the epistemic conditions — in particular about common knowledge — of the players is necessary, in addition to standard tools of game theory. Approaches which deliberately exclude common belief would also need some extra machinery to account for the indeterminacy and ambiguity arising from lack of common knowledge. Also, the issue should be taken into consideration in focal-point search and so forth.

References

Hasida, K.; Nagao, K.; and Miyata, T. 1995. A game-theoretic account of collaboration in communication. In *Proceedings of the First International Conference on Multi-Agent Systems*, 140–147. San Fransisco.

Hasida, K. 1996. Issues in communication game. In *Proceedings of the 16th International Conference on Computational Linguistics*, 531–536. Copenhagen.