Topics in Social Software:
Information in Strategic Situations
(Draft: Chapter 6)

Eric Pacuit
Comments welcome: `epacuit@cs.gc.cuny.edu`

December 11, 2005

Chapter 1

Conclusions

Social software brings together ideas from philosophical logic, game theory and computer science. As is true in many interdisciplinary fields, an important component of social software research is developing a "common language" in which experts from the different fields with diverse backgrounds can compare and contrast their results. The logical systems developed in this thesis are a step in this direction. In particular the frameworks discussed in Chapters 2, 3 and 4 are intended to represent types of situation that are of interest to both game theorists and computer scientists. We briefly summarize the results presented in this thesis.

Chapter 2 introduces the basic formal framework used to represent social interactive situations. We view a social interactive situation as consisting of a collection of sequences of "events" (called histories), where the exact interpretation of an event depends on the application. For example, since we where modeling communication among agents in Chapter 4, an event in Chapter 4 consisted of a (one-way) communication between two agents. Intuitively, each global history (infinite sequence of events) is a possible way the situation could have evolved. At any moment $t \in \mathbb{N}$ there is a finite history and a possibly infinite future. Some of the events are caused by an agent, i.e., an agent can perform a particular action and others are caused by nature (which can be viewed as a special type of agent). In Chapter 2, we show how to start from this basic framework to construct models that have been used by computer scientists (history based knowledge models) to study distributed algorithms and models that have been used by game theorists (extensive games) to study multi-agent social situations.

Chapters 3 and 4 are the main contributions of this thesis. The contribution of Chapter 3 is primarily conceptual. A formal framework for reasoning about actions, knowledge and obligations is described. This framework is shown to naturally capture our intuitions about various deontic dilemmas. Chapter 4 takes on the task of finding a model for multi-agent knowledge and communication. A logical system and semantics is defined (based on the history based structures from Chapter 2) which is shown to be decidable. The analysis in Chapter 5 suggests that a certain level of knowledge is required in order to make the Gibbard-Sattertwaithe Theorem effective.

In Chapter 1, we discussed the three main areas of social software research. This thesis focused on the first of these three areas: mathematical models of social situations. Of course there are a number of issues relevant for developing mathematical models of social situations which we have glossed over or not discussed at all. We point to two of the most relevant and pressing issues.

Logical Omniscience

We have argued that an important part of any formal theory of social procedures, is how knowledge of the individual agents is represented. Of course, without a certain amount of idealization, the mathematics may become too complicated to be of any practical use. As such, we must learn to live with the fact that there may be unrealistic assumptions made about the agents in our mathematical models. One such assumption, called logical omniscience, has been discussed by a number of different authors ([11, 12, 13, 2, 9, 8]). Among other things, logical omniscience implies that an agent knows all the logical consequences of its knowledge (see [11] for a definition of logical omniscience). Perhaps we can live with such an unrealistic assumption. However, as Stalnaker astutely points out, "Any context where an agent engages in reasoning is a context that is distorted by the assumption of deductive omniscience, since reasoning (at least deductive reasoning) is an activity that deductively omniscient agents have no use for. Deliberation, to the extent that it is thought of as a rational process of figuring out what one should do given one's priorities and expectations is an activity that is unnecessary for the deductively omniscient." [12]. This raises some very important questions for the social software scientist. In particular, what exactly is the role that epistemic logic plays in the analysis of social procedures? If we hope to use epistemic logic to show that in a given situation each agent has enough information in order to follow the rules of some social procedure, then clearly the assumption of deductive closure is much too strong.

We briefly comment on some solutions to the logical omniscience problem that have been offered in the literature. The first is to consider logical systems without the K axiom $(K_i(\phi \to \psi) \to K_i\phi \to K_\psi)$. These so-called classical systems of modal logic are interpreted in *neighborhood models*, or Scott-Montague models. Essentially the idea is to shift from thinking of knowledge as being derived from an inaccessibility relation a la Hintikka to being explicitly described as part of the model, i.e. at each world the neighborhood function returns the set of propositions that are known at that world. The textbook [5] has a discussion on these models for propositional modal logic and [1] has a discussion of these models in the context of first-order modal logic. Perhaps the most promising solution to the logical omniscience problem relevant for the analysis of social procedures was offered by Rohit Parikh in [11]. Parikh introduces the notion of *behavioral knowledge*. Say that agent i *b-knows* a formula ϕ if there are three mutually incompatibly actions a, b, c such that i does a only if ϕ is true, does b only if ϕ is false, and i has just done a (there are no restrictions on c). The definition of knowledge in Chapter 2 for history based frames has some of the flavor of behavioral knowledge. In history based frames, knowledge is derived from the agent's observation of the events that have taken place which in turn is based on the behavour of the agents. However, much more can be said, but this is a topic for future research. Finally, Fitting and Artemov have recently been pursuing a different line of reasoning. The main idea is to replace formulas of the form $K_i\phi$ with formulas of the form $t\!:\!\phi$, which are intended to mean that t is a "justification", or provides "evidence", for ϕ. This idea of making modalities explicit was introduced by Artemov in [3] to provide a logic of explicit proofs. See [9, 10, 2] for more on this topic.

Empirical Studies

It is well-known that human agents do not necessarily behave as implied by the models we have described in this thesis. To what extent these models have captured social interactions among human agents is still very much an open problem. Solving this problem requires collaboration with psychologists to design experiments that can test our hypotheses. In fact, there are a number of experiments that have already been conducted that may be of interest to social software researchers. As an example, we discuss one experiment by MIT psychologist Alex Bavelas.

In the 1950s, Alex Bavelas ran a series of experiments on the effects of group structure and communication on task performance - with surprising results. In one of them, which he described to the Cybernetics group at the Interdisciplinary Macy conferences [7, 4] held in New York City between 1946 and 1953, participants were asked to pick a number from the range 0-5, to write it down on a piece of paper and to give it to the experiment moderator. The five participants in the group were supposed to generate guesses so that their total added up to 17. Participants were not allowed to communicate with each other and were not told what the other participants had guessed. The experiments were conducted in two different ways. In one, participants were told whether their guesses had resulted in success or not, without telling them what their total was. So, if in fact the numbers did not add up to 17, the participants were told ‘sorry, try again’ and had to guess again. In the other kind of experiment, the participants were told whether their guess undershot or overshot i.e., whether they had guessed too high or too low (the actual total was also announced). The second experiment then, gave the participants ‘more information’ about their combined guess.

Bavelas reported that the groups always took longer to converge on the correct answer when playing the second, ‘more informative’, kind of game. Prima facie these results are surprising; the second kind of game is ostensibly more informative; an announcement has been made, which is public and the contents of which are common knowledge. Why then do the players take longer to arrive at a correct combined guess in the second form of the game? Note that if a single agent is trying to guess a particular number, then of course the more informative response of the referee will cause the agent to guess the number at a faster pace.

The setup of the Bavelas experiments is reminiscent of many of the epistemic puzzles which have recently been the focus of much scrutiny. In particular, the well-known muddy children puzzle[1] shares a number of features with the Bavelas experiments. In both situations a group of agents are reacting to public announcements. However, in the case of the muddy children the reaction involves an *epistemic update* whereas in the Bavelas experiment the reaction involves a ‘co-ordinated’ response by the group. The difference in the two situations is in what we find puzzling. In the muddy children puzzle, we are surprised that the repeated announcement of seemingly ‘useless’ information (‘some child has mud on its forehead’ and ‘we don’t know

---

[1] See [8] for a discussion of the formal properties of the puzzle.

if we have mud on our forehead') actually increases the agents' information. That is, the announcements in the muddy children puzzle - while seemingly not conveying information - reduce the size of the uncertainty space. In the Bavelas experiments, the announcements - while seemingly reducing the size of the uncertainty space - inhibit the performance of the group on the task at hand. An initial analysis of this experiment is offered in [6].

Finally, we note that the reader may feel somewhat unsatisfied with the conclusions drawn in this chapter. We seem to have raised a number of serious doubts about the applicability of our models. In order to put the reader's mind at ease, we let Johann Von Neumann have the last word. This is particularly fitting as Von Neumann was a very influential figure in both the development of computer science and game theory.

> The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work. — Johann Von Neumann

Bibliography

[1] Horacio Arló-Costa and Eric Pacuit. Classical systems of first-order modal logic. Technical report, Carnegie Mellon University, 2004.

[2] S. Artemov and E. Nogina. Basic epistemic logics with justifications. Technical report, CUNY Graduate Center, 2004.

[3] Sergei Artemov. Logic of proofs. *Annals of Pure and Applied Logic*, 67:29–59, 1994.

[4] Alex Bavelas. Communication patterns in problem-solving groups. In Claude Pias, editor, *Cybernetics — Kybernetik. The Macy-Conferences 1946-1953: Volume 1 Transactions/Protokolle*. Diaphanes, 2003.

[5] Brian Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[6] Samir Chopra and Eric Pacuit. When more is less: The bavelas experiments. 2005.

[7] Jean-Pierre Dupuy. *The Mechanization of Mind*. Princeton University Press, 2000.

[8] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. The MIT Press, 1995.

[9] Melvin Fitting. The logic of proofs, semantically. *Annals of Pure and Applied Logic*, 132:1 – 25, 2005.

[10] Melvin Fitting. The logic of explicit knoweldge. In *Logica 2004 Proceedings*, to appear.

[11] R. Parikh. Knowledge and the problem of logical omniscience. volume ISMIS-87, pages 432 – 439. North Holland, 1987.

[12] R. Stalnaker. The problem of logical omniscience, I. *Synthese*, 89:425 – 440, 1991.

[13] R. Stalnaker. *Context and Content*, chapter The Problem of Logical Omniscience, II. Oxford: Clarendon Press, 1999.