Logic, Agency, and Games, Day Two

Topics covered:

1 *Games* are a microcosm of just about every topic in philosophical logic: action, preference, knowledge, belief, counterfactuals, etc. See the running example from the "Logic in Games" book. What we will see is how logic is a natural fit with games, both static logics and dynamic logics in the sense of yesterday. Eventually we will also discuss how games can be used to analyze logic.

2 One new aspect that games add over and above the 'local dynamics' analyzed yesterday: the passage of *time*, and the new phenomena coming out of that. Example of a longer-term scenario: Muddy Children. Art of modeling still under discussion here: larger or smaller models for this scenario, ours were exponential, but polynomial in terms of number of children suffices. The Father's announcement 'breaks the symmetry'. Open technical problem (related to the nature of 'possible worlds' in our models): what is a natural graph-theoretic notion of complexity (not just counting the number of worlds) that goes down when we update with new information?

3 The Muddy Children follow a conversational *program* involving atomic announcement actions, and the usual operators of sequential composition, WHILE DO, and IF THEN ELSE. This seems realistic, but the logic of PAL plus such program constructions is undecidable, non-axiomatizable, and in fact Π^1_1-complete (the same complexity as true second order arithmetic), shown by Miller & Moss (Bloomington) in 2004. Does this mean that the logic of planned conversation is incredibly complex, or that we have we missed something in our modeling? Or: is complexity of the logical theory of an activity quite distinct from the complexity of that activity itself?

4 Possible way-out: we always live in temporal processes of conversation or inquiry where not all theoretically possible actions (say, announcements of true facts) are available at all moments. This is regulated by so-called *protocols*. Many famous puzzles in probability or epistemology are underspecified since we are not told the temporal protocol (which is why there is such a flourishing inconclusive literature on them). Games are also protocols, since usually, we do not get the full branching tree of all possible moves at each point, but only a selection of available histories. If we make protocols part of the semantic setting, then the old PAL must be modified. In particular, its seemingly innocuous base axiom $\langle !\varphi\rangle p \leftrightarrow (\varphi \wedge p)$ now only keeps one implication valid, namely, $\langle !\varphi\rangle p \rightarrow (\varphi \wedge p)$. As a result, the dynamic language no longer reduces to the static epistemic base logic, but to an extension with irreducible 'procedural atoms' $\langle !\varphi\rangle T$. Protocols are still underappreciated in many areas, but they make a lot of sense logically, and help connect dynamic-epistemic logic and temporal-epistemic logics (paper attached).

5 Games also deeply involve *preference*. This can be treated with binary relations and an ordinary modal logic (see the course materials). The interplay of information and preference pervades agency, decision theory, games, but also natural language. If I ask you a question, I indicate that (a) I do not know the answer, (b) I think that you may know the answer, but also (c) I want to know the answer. And (at least MIT-style) wanting φ is a complex notion of preferring the most plausible φ–worlds (according to my beliefs) to the most plausible $\neg\varphi$-worlds. Language and agency may be about aligning preferences and goals as much as about aligning information.

6 Extensive games viewed as *pure action structures* naturally support logics that describe available actions, strategies, and strategic equilibria. The logics for that are basic modal logic and dynamic logic for programs. Moreover, as we saw with Zermelo's Theorem (yes, the same) on the determinacy of finite-horizon two-player zero-sum games: the modal mu-calculus with recursive definitions fits well, with a typical formula like the syntax of the 'coloring algorithm':

$$WIN_E \leftrightarrow ((end \wedge win_E) \vee (turn_E \wedge \langle move\rangle\, WIN_E) \vee (turn_A \wedge [move]\, WIN_E).$$

General reason: game-theoretic equilibrium is very congenial to *fixed-point logics*.

To be continued on Wednesday with much more game structure.