Logic, Agency, and Games, Day Three

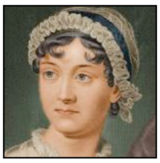
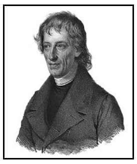
Topic flow:

1 As we saw yesterday, the action structure of extensive games fits a logic of trees, and a modal logic (in several variants) of *actions* with operators [a], [b] does well in capturing basic patterns.

2 However, many games have more structure than mere moves, such as epistemic uncertainty links for players who do not know exactly where they are in games of *imperfect information*. This calls for a bimodal language with also K-operators. Well-known philosophical distinctions then crop up. Example: in imperfect information games, a player may know that (s)he has a winning move ('de dicto') without knowing of any move that it is winning ('de re'). Reasoning principles then express assumptions about players. E.g., $K[a]\varphi \rightarrow [a]K\varphi$ corresponds to a form of Perfect Recall: technically, a commuting grid diagram between actions and uncertainty links (see course materials for details). (This implication can fail for knowledge-impairing actions.)

3 Combining logics. Even when component logics are simple (say, decidable, like for action and knowledge), their combinations can be complex (undecidable or even worse). It all depends on the mode of combination, as a lot of experience in modal logic has shown – see attached paper. (Aside for analytical philosophers: thinking in the opposite direction, 'analysis' of an issue into simple components does not make the whole simple, if the mode of analysis itself is complex.) E.g., some natural logics of knowledge and action are undecidable. Slogan in modal logic: logics of *trees* are simple, logics of *grids* are complex. (The reason: grid structure in its models often allows a logic to encode complex computations or complex geometrical tiling problems.)

4 Basic topic across all sciences (and in philosophy), and especially beautifully, in theories of computation. *Invariances*: when are two given (representations of) processes 'the same'? Different answers represent different 'zoom levels' of detail, and fundamentally (Helmholtz), these levels come with matching *languages*, less or more expressive. Modal logic fits a view of games where local choices for players and local properties matter. Some of my *19th century* heroes:


5 But, using the example of Distribution, we also looked at the rougher level of *powers* for players, i.e., the sets of outcomes they can force by playing one of their strategies. The matching language here, describing players' powers, is modal *neighborhood* logic, weaker than normal modal logic (larger model class), but quite en vogue today for many reasons.

6 Another bimodal combination: modal action logic on game plus modalities for *preference*. As shown in the course material, this can express basic properties of reasoning, e.g., for backward induction (BI). General current trend here: deontic logic as a high-zoom level for games.

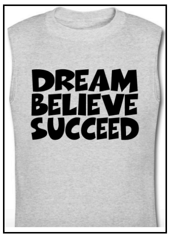
7 For you to ponder. Applying Distribution to our initial example of a game with preference gave two games where players have the same powers, but different BI-outcomes. Open problem: find a good bimodal logic for powers plus preferences that works in games, and axiomatize it.

8 Many more results in this vein can be found in the course notes.

9 Then we switched perspective, from game trees as a sort of silent records of what has taken place, or what could take place, leaving the actual scenario implicit, to dynamic logics of agency: what players actually do. Three phases (in theory, in practice it can be intertwined): pre-game deliberation, in-game action and responses to surprises, post-game rationalization ('spin', important to what we take to be the message learnt when we engage in further game play).

10 We gave only one case study: *Backward Induction* as a pre-game deliberation procedure forming expectations. Scenario 1: Iterated announcement of hard information that all players are rational, that is, never playing moves that are strictly dominated for them. This limit procedure (reminiscent of Muddy Children) always ends in a fixed-point that consists of just the actual BI play, and there is common knowledge of rationality among the players. Scenario 2: Iterated soft announcements that all players are *rational-in-beliefs*. This always ends in a plausibility order where all players are rational, and this fact is even common belief. The final plausibility order encodes exactly the BI-strategy, including its counterfactual moves that are not actually reached. (These results are quite technical, I attach a paper with more details.)

11 Aside, we are not endorsing rationality, we are merely identifying it as the logical form underlying one standard game-theoretic solution method. Once we see that form, we can even use a logical perspective to search for alternatives. E.g., rationality as above is much like classical decision-theoretic/economic maxims like: choose the best action available given your beliefs and preferences. However, on campus I saw student walking around with the following T-shirt:



Maybe there is an alternative decision theory brewing at Rutgers?

12 General issues: limit procedures can be 'self-fulfilling' (like our rationality announcements, common knowledge in the limit) or self-refuting (false in the limit, like the iterated ignorance statement for the muddy children). Both limit types occur widely, witness game-theoretic papers with titles like "We Can't Disagree Forever" (the limit of communicating disagreement is agreement). At present, we do not understand the logical features (formula announced + model structure) very well that lead to one kind of behavior or the other. Moreover, here is another complication: with iterated soft announcements, fixed-points are not guaranteed: there can also be infinite oscillations (nice results on cycling conditional beliefs in learning theory discovered by Baltag and Smets). This leads to analogies with dynamical systems that we may get to later.

13 Resulting program: Logic + Game Theory (+ Computer Science) = *Theory of Play*. Many issues arise and open problems emerge in the latter area. See the course material on challenges like the best logical analysis of Forward Induction: game play when expectations have been defeated, and the past of what has happened (say, what I have observed of your actual behavior) matters. My favorite issue: *player diversity*. Usually, we do not know the type of player we are up against: less or more memory than us, less or more cooperative, etc. This affects everything I have talked about, even the fundamental notion of game equivalence (when are two games equivalent for what type of player?) Game theory has a sort of bulldozer for dealing with this ('type spaces'), our logics provide finer tools, but there is nothing like a systematic theory of agent diversity, and thinking its consequences through. (I may return to Theory of Play on our final day.)