Logic, Agency, and Games, Day One

Topics covered:

1 General introduction: variety of informational tasks that work together in natural scenarios in daily life (but also in science or other expert practices): inference, observation, questions-answers (communication). Program: describe these activities uniformly using logical methods, suitably extended. Basic features to be accounted for: theses activities tend to be dynamic (information changes) and social (agent interaction is important for many, if not all of them).

2 Standard example: the Three Cards. Information about other people's information is essential. Model: graph structure. Successive events update the initial model until we read off what agents know in the final information state.

3 For you to think about:
* *Modeling as an intermediate layer.* Finding such models for intuitive scenarios is a bit of an art. It is also crucial to 'applying logic', since the way in which we model an intuitively given scenario (say, in language) determines how a given logical system is supposed to fit with the facts. This is often neglected: competing logical systems do not confront reality directly, but via a layer of modeling, and the same is true for most applied formal frameworks anywhere. (Notice BTW that what we did *not* do was blindly 'translate' the natural language sentences of the Cards scenario.)
* *Possible worlds.* In the diagram, the possible worlds are the points, not some lush total worlds. So, the 'worlds' terminology is largely nostalgic: say, the way we call the males in our classroom gentlemen (even though, sadly, they do not dance the menuet, or settle their differences through duels). BUT, note that the possible worlds are not just points plus valuations for proposition letters. The identity of a point is determined by what local information it carries plus its connections to *other points*, so things are pretty recursive and delicate (say, like in category theory: an object is determined by its local properties plus how it interacts with other objects).

4 Static language for describing knowledge at snapshots: epistemic logic. Standard syntax, standard semantics in models $\boldsymbol{M} = (W\ \{\sim_i\}_i, V)$: $K_i\varphi$ is true at world s iff φ is true in all worlds that are indistinguishable from s for agent i. Formulas can describe key features of the cards scenario (and much more): especially, iterations with different modalities K_i, K_j express social features. Logic: S5 for each separate agent (the usual idealization for the modality K, which we take to stand for the 'semantic information' that the agent has). There are no significant valid laws relating different agents: say, $K_1K_2\varphi \rightarrow K_2K_1\varphi$ is invalid. Reason: all significant informational relationships must come from agents having special means of communication, which requires group or network structure (a big current research topic in epistemic logic).

5 For you to think about:
* *Collective agents.* Agents do not just interact, they form groups that are new actors. Epistemic logic has some theory on notions of group knowledge such as common knowledge and distributed knowledge, and these, too, have been axiomatized (see the course materials).
* *Real knowledge* in the philosophers' sense. The K modality is not meant to represent real knowledge, but it is still an ingredient in the panorama of proposed more delicate analyses of knowledge in modern philosophy (this is a topic of many current Ph.D. dissertations)

6 Now for the crucial step in the above program: make dynamics explicit in a logical system. Pilot system: *logic of public announcement* PAL (or public observation, PAL is really a bit of a misnomer). Misleading in some ways, but we need to start with a 'proof of concept'. Syntax: now also action expressions $!\varphi$ for updating with the true information that φ is the case, viewed semantically as model change: the current model $(\boldsymbol{M}, s)$ changes into $(\boldsymbol{M}|\varphi, s)$. (Pictures help: see the blackboard.) Updates are non-trivial: e.g., getting new information can change epistemic information about what agents know or do not know (although it cannot change base facts).

To describe this, add a dynamic modality $[!\varphi]\psi$: after the update with φ, ψ is the case. In this notation, e.g., $[!p]Kp$ is valid, but we do not have $[!\varphi]K\varphi$ valid for arbitrary epistemic formulas φ. There is a complete logic for this (see the course material), but most important in its design (and that of all logics of this sort) is one thing: to identify the crucial mechanics of the system, namely, the basic recursion equation of the dynamical system of announcement, if you wish. The single step that drives the system is what agents know after an update:

$$[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$$

This system can deal with the above puzzle, and a large variety of communicative scenarios.

7 For you to think about:
* *Intercultural*. A PAL formula like $[!\varphi]K\psi$ is a bit like a physical artifact such as our glasses, standard objects that merge ideas from many world cultures. The *!* idea is from linguistic speech act theory, the *K* modality is from philosophy, the 'effect' modality *[]* is from computer science.
* *Mutual recursion.* The syntax looks simple but it can generate very complex patterns since *!* makes actions out of formulas, but *[]* makes formulas out of actions. To see this, look up the nice but nested law reducing repeated announcements $[!\varphi][!\psi]\alpha$ to single ones.
* Mathematics behind these systems: [modal] *logics of model change*, a much larger current trend. We now have operators that look inside other models for their evaluation in a given model. (Actually, this fits with old ideas, e.g., from Situation Theory, that many useful inferences consist in using facts from one situation to learn something about *another* situation.)
* *Reduction*. If you read the text, you will see that the axioms of PAL reduce every formula in the dynamic language to one in the underlying static EL language. This explains the completeness (and decidability) of the system. However, reduction is not an inescapable feature: see Tuesday.

8 *Belief and learning*. Importance of belief for our actions (more than knowledge). Beliefs can be wrong: this does not make them less valuable than knowledge, but rather a source of creativity. We correct ourselves when beliefs turn out wrong, that is the basis of learning.

9 Good model for belief: endow epistemic models with a *plausibility ordering* of worlds. The agent believes that φ if φ holds, not in all epistemically accessible worlds, but only in the *most plausible* ones among these. But we need more static plausibility-based notions than plain belief, and taking a dynamic perspective is in fact a good way of finding these.

10 *Hard information* as above: $!\varphi$ uses world elimination. If we now look for a recursion law, we find that we need *conditional belief,* a static notion that 'pre-encodes' belief changes:

$$[!\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[!\varphi]\psi))$$

The laws of static conditional belief are like those of the minimal conditional logic over pre-orders, Lewis-Veltman style. Of course, to close our system, we now also need a recursion law for the new conditional beliefs that agents have after update: see the course material. This setting has many non-trivial scenarios, such as Cheating With The Truth. In the above, if I know a fact, and I get new true information, then I still know that fact. However, if I believe a fact, then new true information may lead to me to switch to a wrong belief (see the course material). Now the dynamics and statics start interacting again. What about a stronger notion of belief that *would* remain stable under getting true new information? The answer is 'safe belief' $SB\varphi$ (many authors have proposed this, in philosophy and CS): φ has to be true in all worlds that are at least as plausible as the current one. *SB* is the ordinary universal S4-modality for plausibility, intermediate in strength between *B* and *K*. It is often used to simplify our logics technically.

11 For you to think about:
* *Plausibility order* may be seen as representing our accumulated information/experience so far, at a high zoom level, but it is also the basis for our future expectations.
* *Harmony*. Static notions can suggests dynamic ones, dynamic notions suggest new static ones.

* *Repertoire*. We are pretty sophisticated information processors. What is the natural repertoire of notions governing agents like us, beyond knowledge and belief, inherited from the tradition?
* *Probability*. Plausibility is qualitative: what about quantitative degrees of belief? Yes, see other courses, but the dynamic-epistemic methodology fits well also with a probabilistic base system.

12 *Soft information*, where we do not eliminate worlds, since we take the update just to be 'in favor' of φ. This is done via ordering change, of which there are many varieties: see the course material. One of these, radical upgrade $\Uparrow\varphi$, puts all φ–worlds on top in the new plausibility order. We will use this later in our analysis of solution scenarios for games. Again, there are complete dynamic logics for this that turn on suitable recursion laws for new beliefs (conditional beliefs, safe beliefs) after incoming soft information, but we refer to the course material for details. As an update mechanism, ordering change is much more general than world elimination, and it covers a much wider range of phenomena, including also changes in agents' preferences, what they think relevant, their goals, or the current issues that drive conversation or inquiry.

13 *Groups*. Well, I did not get to this. Hopefully later on in the course.

PS New Brunswick. Two pictures included: William the Silent (the founder of the independent Netherlands in the 16th century, who of course did speak when it mattered), Wedding Ring Sale (are social relationships breaking down, or booming in this town?).

PPS Hector Malot's classic book that had many hardened Dutch logicians crying in their youth: https://en.wikipedia.org/wiki/Sans_Famille