

## Modeling Reasoning in a Social Setting

Johan van Benthem, Amsterdam & Stanford

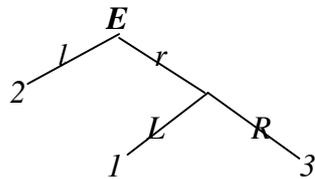
May 2012

*Abstract* In this survey and position paper, we discuss some issues in logical modeling of interactive behavior. We draw together a number of lines in current logics for social action, emphasizing uses of ‘small models’ rather than complex spaces.

### 1 Introduction

Reasoning in social settings is something we do all the time, as we are making sense of what others do in interaction with ourselves. This practice is subtle. With other minds and bodies around, it is easy to be wrong in one’s expectations. Here are some simple pieces of reasoning that logics of social scenarios should be able to deal with.

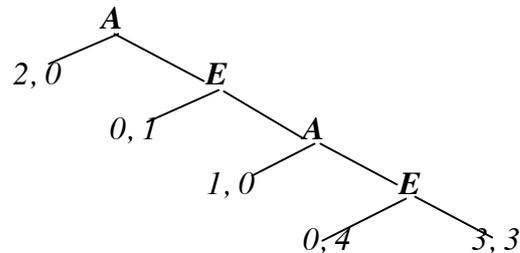
*Example 1* Let agent  $E$  have actions *left*, *right*, while the second stage to the right is a fork that is not under her control. The numbers at the leaves record utilities for her:



Let us assume that  $E$  does not know what will happen when she chooses  $r$ , though she may have expectations – for simplicity, in the form of unique continuations that she considers most plausible. We can make sense of different behavior here. If  $E$  in fact chooses action  $l$  with outcome 2, we would think that she considers  $R$  less plausible at the second stage, since outcome 2 is better than 1. But if she were in fact to choose  $r$ , we would probably think that she considers  $R$ , rather than  $L$ , most plausible, aiming for outcome 3. But we can also complicate matters. Suppose that in fact  $E$  considers  $R$  *less* plausible than  $L$  for the second stage. Could not she still choose  $r$ , hoping to exploit the fact that some other player with some influence over outcomes will now be misled, like us, into thinking that she considers  $R$  the more plausible outcome? ■

It is easy to come up with cases in daily life where such considerations about behavior are crucial, even though they need not be fully explicit in one's mind. One systematic approach are the solution procedures of game theory. But similar issues persist:

*Example 2* Let two players *A*, *E* play the following simple extensive game, with their pay-offs at the endpoints marked in the order (*value for A*, *value for E*):



Here, the well-known solution algorithm of *Backward Induction* recommends moving *left* at each stage of the game. Thus, strikingly, player *A* will opt out at the start. But there is another plausible line of reasoning. If *A* goes right, then player *E* can only make sense of this by assuming that *A* expects to get the outcome 3, 3.<sup>1</sup> But this will change *E*'s behavior: she will now play *right* at her first turn, and if *A* in fact expected 3, 3, he will choose *right* once more. What happens at the end is up for grabs. You may think that *E* will go left, callously disregarding *A* – if this is a competitive game. But labels like competitive or cooperative are not all that useful in social scenarios. Many real-life settings are a bit of both: think of academic life. We want logics that can analyze scenarios like this without agents knowing which general label applies. ■

In this paper, we will not propose any definitive solution for the above cases. Instead, we see them as exemplars for an area of reasoning that we will explore along the following lines. Any game-like activity can be divided into three meaningful phases: *Before*, *During*, and *After*. We start with *Before*, looking at Backward Induction as a prior method of deliberation creating expectations about a social scenario. This can be done in several different logical formats, each of independent interest, and we point out some interesting open problems. Next, we look at the *During* phase, discussing logics for updating players' information and beliefs as a game proceeds. One natural highlight here is Forward Induction as a form of belief revision in perfect information

---

<sup>1</sup> Of course, *E* can also think of *A*'s move as a *mistake*. Then Backward Induction prevails.

games, but we also discuss the natural issue of observational powers for various types of players in imperfect information games. We will not say much about the final *After* phase of analyzing or rationalizing a game already played, but the logical techniques that we introduce would also work for such scenarios. One more technical theme in logical modeling throughout this paper is the use of ‘small models’ close to simple reasoning patterns, as opposed to huge spaces pre-encoding all possible developments. In addition, we provide a large number of asides that draw the reader’s attention to further issues and open problems in the area. Finally, this tour of the area does not add up to one systematic theory. While we may be seeing the contours of a ‘Theory of Play’ (van Benthem 2011A), this promised land is still beyond the horizon here.

## 2 Backward induction, logic of best action, and beliefs

We start with a typical benchmark in the area of strategic reasoning about social scenarios. Let us analyze the logic of Backward Induction (*BI*) as a sort of advice for ‘best action’, following van Benthem, van Otterloo & Roy 2006.

**Relational strategies** To do so, we generalize the functional notion of strategy to a *subrelation* of the total *move* relation. This will allow us to bring standard logics to bear. Here is how *BI* works then. Mark all nodes of the game as active. Call a move *a* *dominated* if it has a sibling move all of whose reachable endpoints via active nodes are preferred by the current player to all reachable endpoints via *a* itself. Now:

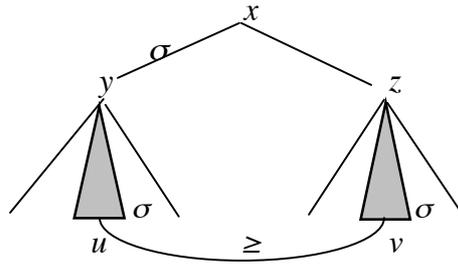
*Working in stages, mark dominated moves as passive, leaving all others active.*

At each stage, reachable endpoints are those reachable via a sequence of moves that are still active at this stage. As can be seen in simple examples, repeating the procedure makes sense: fewer and fewer moves remain active, until a fixed-point is reached.

**Theorem 3** On finite extensive games, the *BI* strategy is the largest subrelation  $\sigma$  of the total *move* relation satisfying the following property for all players *i*:

**RAT** No alternative move for the current player *i* yields outcomes via further play with  $\sigma$  that are all strictly better for *i* than all outcomes resulting from starting at the current move and then playing  $\sigma$  all the way down the tree.

The following picture illustrates more concretely what this says: <sup>2</sup>



The shaded area is the part that can be reached via further play with our strategy.

**The logical form of best action** Backward Induction produces, among all my moves, my best actions. It makes sense to throw away the ladder of the above procedure, trying to capture this style of reasoning directly. One natural candidate is a modal game logic with (i) modalities  $\langle a \rangle \varphi$  for moves as binary relations  $a$  on nodes in the game tree (‘playing  $a$ , one can reach a node where  $\varphi$  holds’), and (ii) a preference modality  $\langle \text{pref}_i \rangle \varphi$  (‘player  $i$  prefers some node where  $\varphi$  holds to the current one’). In terms of this language, a standard modal frame correspondence argument (cf. Blackburn, de Rijke & Venema 2001) then analyzes the logical form of rationality, as a subtle entanglement of action and preference:

*Fact 4* On game trees, RAT corresponds to validity of the following modal axiom:

$$\&_i (\text{turn}_i \wedge \langle \sigma \rangle [\sigma^*](\text{end} \rightarrow p)) \rightarrow [\text{move-}i] \langle \sigma^* \rangle (\text{end} \wedge \langle \text{pref}_i \rangle p).$$

This logic of best action fits in a current trend toward investigating simple top-level modal logics for reasoning about complex underlying normative notions. Still, a complete axiomatization of this system is unknown (cf. van Benthem 2011C). <sup>3</sup>

**Strategies and beliefs** But our analysis also suggests other interesting perspectives on reasoning about action. Intuitively, in the above picture, the shaded moves also correspond to a current *belief* about how the game will proceed. This matches a natural modification of dominance to ‘rationality-in-beliefs’:

*No player plays an action she believes to be worse than another available one.*

<sup>2</sup> This says that players avoid *strictly dominated actions*, if we make the comparison looking recursively at the available further histories computed via the method of Backward Induction. Van Benthem & Gheerbrant 2010 provide a precise definition in standard fixed-point logics.

<sup>3</sup> In particular, it seems that we need additional laws for notions of conditional best action.

A common semantics of beliefs reads them as what is true in all most plausible worlds. We can also view the *BI* algorithm as producing *most plausible* moves or histories. Here is why (Baltag, Smets & Zvesper 2009, van Benthem & Gheerbrant 2010):

*Fact 5* There is a one-one correspondence between strategies as subrelations of the *move* relation and connected plausibility orders over all histories of the game.

Here the plausibility orders  $\leq$  must be ‘tree compatible’, meaning that, for any two nodes  $x, y$  in the game tree, all reachable endpoints below  $x$  precede those below  $y$  in  $\leq$ , or vice versa. We will return later to the logical content of this condition.

***Plausibility models and logic of belief*** Plausibility orders  $\leq$  are a well-known modal semantics for belief in possible world models  $\mathbf{M} = (W, \leq, V)$  with worlds  $W$  and a valuation  $V$  for proposition letters (van Benthem 2007A). For later reference, we show how this would work even with agent- and world-dependent plausibility orders:<sup>4</sup>

*Definition 6* Here are the truth conditions for absolute and conditional belief:

$\mathbf{M}, s \models B_i \phi$  iff  $\mathbf{M}, t \models \phi$  for all worlds  $t$  *minimal* in the ordering  $\lambda xy. \leq_{i,s} xy$

$\mathbf{M}, s \models B_i^* \phi$  iff  $\mathbf{M}, t \models \phi$  for all  $\leq_{i,s}$ -minimal worlds in  $\{u \mid \mathbf{M}, u \models \psi\}$ .

This completes our first, ‘classical’ logical take on the typical mixture of action, preference and belief that drives so much of social behavior.

### 3 Dynamics of prior deliberation

While modal semantics for knowledge and belief usually takes its accessibility relations as given, recent work on dynamic-epistemic logics *constructs* the latter in a process of update and revision. In particular, Backward Induction may be seen as a procedure for deliberation, creating knowledge or belief in a game tree. Two ways have been proposed for making this precise in dynamic logics. We will discuss both.

***Repeated public announcement of rationality*** Here is the driving force for the first procedure. Consider the following statement of rationality for players in a game:

***rat*** No player has chosen a strictly dominated move in coming to the present node.

Some nodes will satisfy this, others may not. So, announcing ***rat*** changes the game:

---

<sup>4</sup> Often, the plausibility order only runs within the set of epistemically accessible worlds.

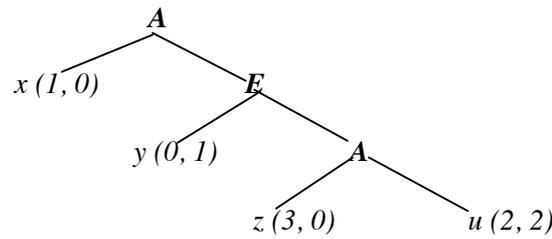
*Definition 7* Public announcement.

Announcements  $! \varphi$  of a true formula  $\varphi$  transform models  $(\mathbf{M}, s)$  with actual world  $s$  into submodels  $(\mathbf{M}/\varphi, s)$  whose domain consists of those worlds in  $\mathbf{M}$  that satisfied  $\varphi$ .

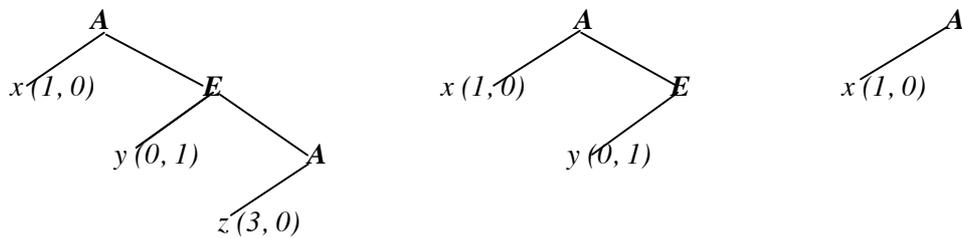
Moreover, crucial for dynamic epistemic changes, in the new model, truth values of *rat* statements may have changed, so announcing it again makes sense.

*Example 8* Solving games by iterated announcement of rationality.

Consider a game with three turns, four branches, and pay-offs in the order  $A, E$ :



The *BI* path emerges step by step. Announcing *rat* once rules out point  $u$  (the only one where *rat* fails), announcing it again rules out  $z$  and the node above it (now points where *rat* fails), and the final stage rules out  $y$  and the node above it. In the resulting game tree, *rat* holds throughout – it has become common knowledge, one might say:



Stable limit models  $(! \varphi, \mathbf{M})^\#$  are always reached in iterated announcement procedures.

In particular, this process captures Backward Induction (van Benthem 2007B):

*Theorem 9* In any game tree  $\mathbf{M}$ , the limit submodel  $(! \mathit{rat}, \mathbf{M})^\#$  produced by iterated announcement of rationality *rat* is the history computed by the *BI* algorithm.

This limit perspective applies in principle to any sort of informational process.<sup>5 6</sup>

<sup>5</sup> The original case in van Benthem 2007B was iterated removal of strictly dominated strategies in strategic games. Pacuit & Roy 2011 analyze other game-theoretic scenarios in this way.

<sup>6</sup> Our notion of rationality in terms of moves differs from that of Aumann 1995, who refers to complete strategies. The latter has been investigated in the logics of Lorini & Moisan 2011.

**Repeated plausibility upgrade with rationality** Backward Induction can also be seen as creating expectations rather than knowledge. Then we need a ‘soft update’ that rearranges plausibility between worlds. A key example is the following operation:

*Definition 10* Radical upgrade.

*Radical upgrade*  $\uparrow\varphi$  makes all  $\varphi$ -worlds best in a current plausibility model, and puts all  $\neg\varphi$ -worlds underneath, while keeping the old ordering inside these two zones.

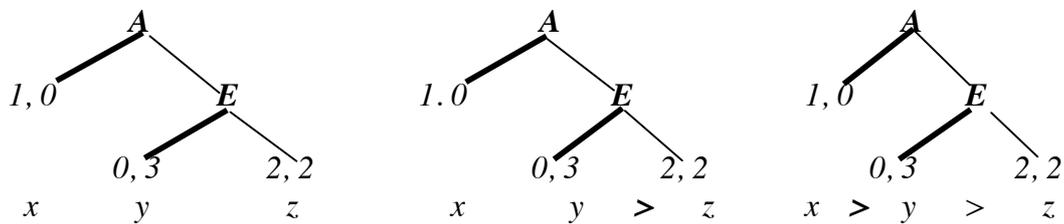
The driving force this time is the usual notion of rationality in games. A move  $x$  for player  $i$  *dominates* its sibling  $y$  *in beliefs* if the *most plausible* end nodes reachable after  $x$  along any path in the whole game tree are all better for  $i$  than all most plausible end nodes reachable in the game after  $y$ . *Rationality-in-beliefs* is then the assertion

$\mathit{rat}^*$  No player plays a move that is dominated in beliefs.

Now we perform a relation change  $\uparrow\mathit{rat}^*$  that is like a radical upgrade:

If  $x$  dominates  $y$  in beliefs, make all end nodes from  $x$  more plausible than those reachable from  $y$ , keeping the old order inside these zones.

This changes plausibility order, and hence the dominance pattern, so iteration makes sense. Here are the stages in our earlier game, with  $x, y, z$  for end nodes or histories:



At first, going right is not dominated in beliefs for  $A$  by going left, so  $\mathit{rat}^*$  only affects  $E$ 's turn, and upgrade makes  $(0, 3)$  more plausible than  $(2, 2)$ . But then, going right is dominated in beliefs, and the next upgrade makes  $A$  going left most plausible. ■

Here is the general result (van Benthem & Gheerbrant 2010 have more background):<sup>7</sup>

*Theorem 11* On finite trees, the Backward Induction strategy is encoded precisely in the plausibility order for end nodes (i.e., histories) created in the limit model  $(\uparrow\mathit{rat}^*, \mathbf{M})^\#$  of iterated radical upgrades with rationality-in-belief.

<sup>7</sup> For more on the general logic of limit behavior, see van Benthem 2007B, Gierasimczuk 2010.

Thus, we have now linked social reasoning to information update and belief revision.

#### **4 Phases of action, styles of reasoning, and modeling choices**

***From deliberation to real-time dynamics of play*** Backward Induction is a powerful way of analyzing social scenarios before they happen. But its style of deliberation is very special: one only looks at the *future* of a game after the current node, the history so far does not count.<sup>8</sup> But, when playing a game, that history is a major source of information. We observe moves that may be significant cues causing changes in our expectations. *BI* has rivals in that setting, as we saw in Example 2. Surprise moves may inform us about the type of player that we are up against. Thus, Stalnaker 1999 argues for an additional parameter in analyzing games: players' policies for *belief revision*. In principle, this new structure might create a large diversity, but some general procedures exist. In particular, if the past can be informative, then we should factor it in – as we did in Example 2, where we interpreted an observed move in the light of what players might have chosen earlier, but did not. A general method staying close to our earlier analyses rationalizing observed behavior in terms of ascribing beliefs to players that make them rational-in-beliefs. In game theory this is known as *Forward Induction* (Battigalli & Siniscalchi 2002, Perea 2011). We will return to this procedure later, in terms of logical models for the belief revision at work here.

***Spaces of hypotheses about others*** But rationalization is not always the preferred method during play. In other scenarios, hypotheses about players can be very different. We may know that we are playing against a finite machine, or an overworked colleague, and to optimize our own strategy, we need to find out how much memory it has available. Van Benthem 2011A discusses a taxonomy of hypotheses about players, needed to tame the explosion of options that arise in a more general Theory of Play.

***Small models versus large models*** These options in thinking about social behavior also have a technical counterpart in a basic point of logical modeling. Game-theorists tie Forward Induction to solution algorithms on strategic games such as iterated removal of weakly dominated strategies (Brandenburger 2007, Perea 2011). The

---

<sup>8</sup> Technically, the *BI* strategy is *invariant for generated submodels*. Another special feature is *uniform*, node-independent plausibilities between histories – even the same for both players.

semantic justification for these procedures involves large ‘type spaces’ encoding a huge set of all possible hypotheses about players’ strategies (Osborne & Rubinstein 1994).<sup>9</sup> While this may work well as a theorist’s perspective, such large models are hard to relate to actual practice of agents. An opposite trend in the logical literature is the use of ‘small models’, staying close to an image that an agent herself might form of the situation she is in. In what follows, we take this small model stance, close to our reasoning scenarios. We will also consider matching static and dynamic logics.

***Complexity jumps: diversity of knowledge*** Small is relative. There are natural jumping points for model complexity, and a typical trigger for these is the fact that knowledge and beliefs of players come in at least three different kinds (van Benthem 2000). At any point in an extensive game of perfect information, players *know the future* in the sense of all possible continuations – and by the same token, they also know the past. This knowledge is based on understanding the game, though this kind of ‘procedural information’ can also be varied. But in games of imperfect information, players also have another kind of knowledge (and ignorance) about precisely *where they are in the game tree*. This time, the source of the information is different, depending on players’ powers of observation and memory. And there is even a third source of information. Players may have knowledge *about other players*, their general types, or their current strategies. This may be based on experience, or having been informed from outside of the game. Our ambition in the sections to follow is to make such distinctions visible, rather than wash them away in supermodels for some uniform sort of knowledge.

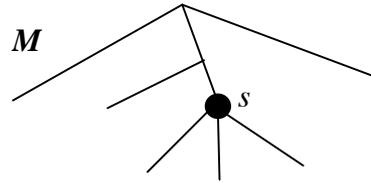
## **5 Dynamic temporal logic of game trees**

We start at the simplest level of analysis, as in Sections 2, 3. Game trees themselves are models of possible actions, and at the same time, they model future knowledge. Much relevant structure can be brought out then by means of a propositional dynamic logic of simple and complex programs. This language can express actions, strategies, counterfactuals, and with one small addition, expectations about the future.

---

<sup>9</sup> The mathematics of the approach is like universal spaces in category theory, or complete-universe models for physical theories. There may be a trade-off here between *simple algorithm* – *complex semantic model* for its correctness. If this balance of complexity holds, then our later simpler semantic models may have to come necessarily with more complex algorithms.

**Pointed game trees as local models** Consider game trees  $M$ , carrying atomic predicates at nodes, with a distinguished point  $s$  that indicates how far play has progressed:



At the actual stage, players know what has happened, and let their behavior depend on what the remaining game looks like, and what happened so far in the larger game.

**Modal language of action and knowledge** Pointed game trees support many logical languages, depending on the reasoning we want to represent (van Benthem 2002). One good formalism is *propositional dynamic logic PDL*, combining propositions and programs. In its simplest version, we have modalities talking about moves

- $\langle a \rangle \varphi$       after performing an  $a$  move,  $\varphi$  holds,
- $\langle a^* \rangle \varphi$       after performing some finite sequence of  $a$  moves,  $\varphi$  holds.

This language can define strategic powers for players in games. But in addition to this action view, there is also the earlier ‘future knowledge’. Adding a general term *move* for the union of all specific moves,  $\langle \text{move}^* \rangle \varphi$  is a description of the modal-epistemic future, running over all positions that can still be reached by further play. Unlike most epistemic logics in game theory, this is *S4* knowledge with reflexive and transitive accessibility, fitting the temporal character of information about the future.<sup>10</sup>

**Looking backward** While the modal logic *PDL* looks forward in time, there is also a case for looking backward. Consider the *BI* strategy as analyzed earlier by announcing the statement *rat* which said that a node is not the result of a mistake, passing through a dominated move. This looks backward and we would add a converse modality

- $\langle a^\circ \rangle \varphi$     there is a node with  $\varphi$  reaching the current node by an  $a$  move

Given the structure of trees, we can then also define an operator  $Y$  for ‘yesterday’, plus propositional constants marking what last move was played (if we are not in the root).

---

<sup>10</sup> *PDL* can also define other notions of knowledge, such as truth in all reachable endpoints. For an analogy with *intuitionistic logic* as a system of information, cf. van Benthem 2009.

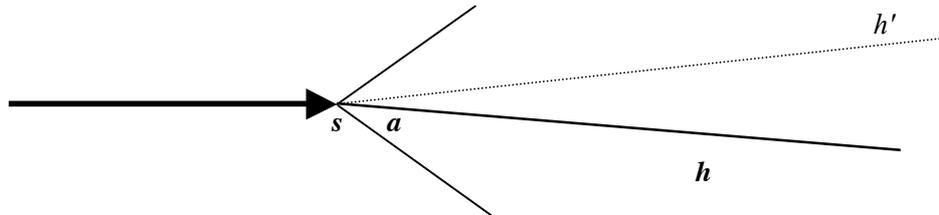
Another new definable expression is *counterfactuals*, since this extended language can talk about what would be the case if we had chose different moves in the past.<sup>11</sup>

**From nodes to histories: temporal logics** Modal languages typically express local properties of *nodes* in a game tree. But when reasoning about long-term social action, one talks about global *histories* that unfold over time. Now *PDL* can still do part of this job. The earlier modality  $\langle \text{move}^* \rangle \varphi$  already acted like a temporal assertion that at some history at some stage,  $\varphi$  will be the case. But we also want to say things like

$\exists G\varphi$  along some future history  $\varphi$  will always be the case.

This is still definable in *PDL* on finite trees, as programs can express the statement ‘until some endpoint,  $\varphi$  is always the case’ (van Benthem 2002). In infinite games, the temporal modality becomes stronger, and we need an extension of *PDL* that can talk about infinite branches, such as the  $\mu$ -calculus.<sup>12</sup> But there are alternatives:

Game trees are also models for a *branching temporal logic* where ‘worlds’ are complete histories of the game. In tree-like models for branching time, a ubiquitous structure, ‘legal histories’  $h$  represent possible evolutions of a given game. At each stage of the game, players are in a node  $s$  on some actual history of events whose past they know, either completely or partially, but whose future is yet to be fully revealed:



This can be described in a language with temporal operators plus ‘branch knowledge’, where  $s \langle a \rangle$  is the partial history arising when event  $a$  occurs after stage  $s$ :

- (a)  $\mathbf{M}, h, s \models F_a \phi$  iff  $s \langle a \rangle$  lies on  $h$  and  $\mathbf{M}, h, s \langle a \rangle \models \phi$
- (b)  $\mathbf{M}, h, s \models P_a \phi$  iff  $s = s' \langle a \rangle$ , and  $\mathbf{M}, h, s' \models \phi$

<sup>11</sup> To formalize the rationality statement *rat* completely, one needs a *hybrid extension* of the modal language referring to specific nodes, plus perhaps other devices. However, there is no duty to formalize everything down to rock-bottom level.

<sup>12</sup> The  $\mu$ -calculus can define recursive game solutions like those for Zermelo’s Theorem.

(c)  $\mathbf{M}, h, s \models \Box_i \phi$  iff  $\mathbf{M}, h', s \models \phi$  for some  $h'$  equal for  $i$  to  $h$  up to stage  $s$ .

In this language, as moves are played publicly, players' knowledge adapts:

*Fact 12* The following principle is valid:  $F_a \Box_i \phi \Leftrightarrow (F_a T \ \& \ \Box_i F_a \phi)$ .

Here ' $T$ ' stands for a tautology. Indices of evaluation are history-point pairs  $(h, s)$  rather than local stages – but we will mainly stick with the latter in this paper.

***Update actions during play*** A game tree is a stage recording all that can happen. But during real play of a game, things change: players observe moves, often publicly. In addition, in real social scenarios, other sorts of information come in, where the simplest case is a public announcement about the further course of history. Examples are promises (van Benthem 2006), announcements of players' plans (van Otterloo, van der Hoek & Wooldridge 2003) or intentions (Roy 2008). How can we make this dynamics explicit in a richer logical system that does not just represent reasoning about a static situation, but also the effects of events and information flow?

*Definition 13* Basic updates.

Events  $e$  that happen publicly lead to simple *public observations*  $!e$ :

$(\mathbf{M}, s) \Rightarrow (\mathbf{M}, s^e)$ : the distinguished point  $s$  of the current model moves  
to  $s^e$ : the next point in the game tree after the event  $e$  has happened.<sup>13</sup>

*Public announcements*  $!\varphi$  are defined using world elimination from  $(\mathbf{M}, s)$  to  $(\mathbf{M}/\varphi, s)$  as in Definition 7, now to the above tree models with stages serving as worlds.

***Reduction axioms for dynamic logics*** We can analyze the effects of these basic actions explicitly on top of our static game languages, by adding the usual modalities  $[!e]$  and  $[\!\varphi]$ . For the first operator, we need not do anything special:

*Fact 14*  $[!e]\varphi$  is equivalent to  $[e]\varphi$ .<sup>14</sup>

The second operator yields to standard techniques, via recursion axioms:

*Theorem 15* The logic of public announcement in game trees is axiomatizable.

---

<sup>13</sup> Here we assume, as is often done, that moves are fine-grained enough to be unique.

<sup>14</sup> Compare the wide-spread dynamic-looking labeled-arrow presentation for labeled transition systems in computer science, which matches a standard modal operator language after all.

The recursion axioms for atoms and Boolean operators are as usual. Here is the law for the action modality. For convenience, we use existential versions:

$$\langle !\varphi \rangle \langle a \rangle \psi \leftrightarrow (\varphi \wedge \langle a \rangle \langle !\varphi \rangle \psi)$$

Interestingly, the case of iteration (and hence also of future knowledge)  $\langle !\varphi \rangle \langle a^* \rangle \psi$  is a bit less obvious, since we now need to make sure that we run along  $\varphi$ -points only. To state the right law, we must move to *PDL with test*:

$$\langle !\varphi \rangle \langle a^* \rangle \psi \leftrightarrow (\varphi \wedge \langle (? \varphi; a)^* \rangle \langle !\varphi \rangle \psi)^{15}$$

But then, we really need to show that *PDL* as a whole has reduction laws for public announcement. The following technical folklore result shows this:

*Fact 16* The logic *PDL* is closed under relativization.

In particular, we have the following principle generalizing the above:

$$\langle !\varphi \rangle \langle \pi \rangle \psi \leftrightarrow (\varphi \wedge \langle \pi / \varphi \rangle \langle !\varphi \rangle \psi)$$

Here  $/\varphi$  is a recursive operation on *PDL*-programs  $\pi$  surrounding every occurrence of an atomic move  $a$  with tests to obtain  $? \varphi ; a ; ? \varphi$ .

As noted earlier, *PDL* has a further virtue of explicitly defining *strategies* as programs. Thus, we also obtain recursion laws for strategy modalities (cf. van Benthem 2010 for details). As we saw with the *BI* strategy, we need converse action modalities to define rationality, but Theorem 15 is easily extended to *PDL* with a converse operator.

This same style of analysis applies to the above-mentioned temporal languages. As an example, here is a reduction law for the earlier additional branch modality  $\exists G$ :

$$\langle !\varphi \rangle \exists G \psi \leftrightarrow (\varphi \wedge \exists G \langle !\varphi \rangle \psi)$$

This is just one instance of connecting public announcement with temporal logics.<sup>16</sup>

---

<sup>15</sup> This is like the ‘conditional common knowledge’ of van Benthem, van Eijck & Kooi 2006.

<sup>16</sup> Van Benthem 2010 discusses pure temporal logic analogues for the recursion laws of public announcement, including the equivalence  $F_a \langle \rangle \phi \leftrightarrow (F_a T \ \& \ \langle \rangle F_a \phi)$ .

Finally, our update analysis also applies to beliefs as subrelations of the *move* relation. The recursion axioms of van Benthem 2007A then hold for public announcement.<sup>17</sup>

Game trees are an appealing first model for social action. Simple dynamic logics of actions and information change describe play when the game is common knowledge, while beliefs of players match subsets of the move relation. The result is a mix of ideas from dynamic and temporal logic (Stirling 1993, van Benthem & Pacuit 2006).

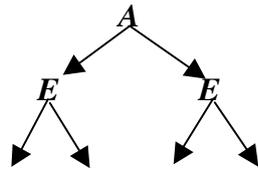
## 6 Procedural information and epistemic forests

Now that we know what game trees can do for us, we go to more complex models.

**Lifting to forest models** Suppose we do not know what strategies are played by others, but we have some prior idea. Then tree models may not suffice (van Benthem 2011C):

*Example 17* Strategic uncertainty.

In the following game, let *A* know that *E* will play the same move throughout:



Then all four histories are still possible. But *A* only considers two *future trees*, viz.



In longer games, this difference in modeling can be very important. Observing only one move by *E* will tell *A* exactly what *E*'s strategy will be in the whole game. ■

In addition to uncertainty about actions or beliefs of other players, there may also be lack of information about which game is being played, even just for one agent. To deal with both, we need semantic structures containing different game trees, linked by epistemic uncertainties between various views of the game. These are the ‘epistemic forests’ of van Benthem, Gerbrandy, Hoshi & Pacuit 2009, and they are essentially also the models for the epistemic-temporal logic of Parikh & Ramanujam 2003:

---

<sup>17</sup> This involves some complexities. *Best* moves computed by Backward Induction in a game tree *M* need no longer be best when we run *BI* once more in a definable submodel *M|φ*.

*Definition 18* Epistemic forest models.

An *epistemic forest model*  $\mathbf{M}$  is a family of finite sequences  $h$  of events taken from some given set  $E$ , related by extension with further events (the possible moves), and also by epistemic accessibility relations  $\sim_i$  for agents  $i$ . There is also a valuation map  $V$  from proposition letters to those finite sequences  $h$  where they hold.

The finite histories encode possible stages of a process that goes from available sequences to longer ones. This need not form a single tree, as there may be different (views of) processes under consideration. The epistemic uncertainty links make agents uncertain where they are in which process – which can be for many different reasons.

These structures have more generality than we just asked for. So far, inside games, we do not want uncertainties about the stage one is in. We are still talking about perfect information games, but we may not know exactly which one, or how it is played.

*Fact 19* Epistemic forest models for perfect information games satisfy the following condition: if  $h \sim h'$  (for any player  $i$ ) and  $h (\text{move} \cup \text{move}^\omega)^* h'$ , then  $h = h'$ .

***Static logics of action and a new kind of knowledge*** Epistemic forest models support several static languages as in Section 5. Modal languages include the earlier *PDL*, now extended with epistemic operators. That is, suppressing agent indices for convenience, there will then be a primitive knowledge modality  $K\varphi$  of a usual kind:

$$\mathbf{M}, h \models K\varphi \text{ iff } \mathbf{M}, h' \models \varphi \text{ for all histories } h' \text{ with } h \sim h'.$$

So now, two kinds of knowledge are represented in the system:

- (i) the future knowledge expressed by the earlier  $[\text{move}^*]$
- (ii) standard  $K$ -type knowledge, perhaps from other sources.

The logic will show us how these can interact, and we will see an example below.<sup>18</sup> But one can also use temporal languages interpreting formulas at history-point pairs  $(h, t)$  as before. The modal and temporal approaches to epistemic forest structure seem close, though I am not aware of any definitive analysis of their precise connections.

---

<sup>18</sup> The agent condition of Fact 18 can typically be defined in such a modal logic of action and uncertainty, using *nominals* to denote unique finite histories  $h$ .

*Dynamic logics of update* Likewise, the dynamics of earlier sections makes equal sense here: forest models can change under observation of moves as a game is played, and under public announcements of further information from other reliable sources.

*Public occurrences* of moves work as before. Take epistemic forests with a distinguished history  $h$  encoding what has happened so far. Event  $e$  shifts this history  $h$  to  $h \cdot e$ . As before, the logic of the dynamic modality  $\langle !e \rangle$  is just that of the move modality  $\langle e \rangle$ . But more can be said. If observing moves is the only way in which information arrives, then the epistemic forest must satisfy conditions on the interplay between moves and uncertainty links, encoding what agents know. We started from uncertainty links between the roots of some trees. How can these uncertainty links evolve?

*Definition 20* Perfect Recall and No Miracles.

An agent  $i$  satisfies *Perfect Recall* if uncertainty after a move  $e$  can only come from earlier uncertainty:  $\forall xyz: ((xR_e y \wedge y \sim_i z) \rightarrow \exists u: ((x \sim_i u \wedge uR_e z))$ . An agent  $i$  satisfies *No Miracles* if uncertainty before a move  $e$  must persist after that same move, links can only be broken by different observations:  $\forall xyuz: ((x \sim_i y \wedge xR_e z \wedge yR_e u) \rightarrow z \sim_i u)$ . ■

These properties are also all that we need:

*Theorem 21* An epistemic forest model is produced (modulo isomorphism) by successive moves in a game of perfect information whose structure is not common knowledge iff all agents satisfy Perfect Recall and No Miracles.

The proof of this result is essentially a special case of the proof for the representation theorem in van Benthem, Gerbrandy, Hoshi & Pacuit 2009. The two stated properties also have interesting consequences about agents, that can be formally derived:

*Fact 22* Perfect Recall implies that agents know their complete past history.

It also implies the condition stated in Fact 19.

A similar analysis can be given for *public announcements*  $! \varphi$  of relevant facts. These change an epistemic forest model in the same style as before.

*Theorem 23* The dynamic logic of public announcement on epistemic forest models is completely axiomatizable.

The only new feature is the recursion law for the new knowledge operator. But it is easy to see that it is just the standard one of public announcement logic:

$$\langle !\varphi \rangle \langle K \rangle \psi \leftrightarrow (\varphi \wedge \langle K \rangle \langle !\varphi \rangle \psi)$$

**Discussion: internalized announcements** In the preceding dynamics, epistemic forest models change. In particular, public announcements simplify the initial forest, perhaps even to just one tree. However, this is not the usual view of these models as Grand Stage accounts of every event that can happen in a given information scenario. On the latter view, public announcements become ‘internalized’ to events that can happen inside the game, already marked inside the forest model (cf. the ‘protocol models’ of Hoshi 2009). We will not undertake a comparison between the two views.

We have shown how epistemic forest models can be updated in two ways: with new facts about where we stand in the action structure of a game, but also with global procedural information about the structure of the game as a whole. Our models support a variety of representation languages, of which we have mentioned a few.<sup>19</sup> The style of analysis given here also works for our next class of game models, where knowledge and information play in yet another way, while updates can be much more delicate.

**Coda: related frameworks** Many related logical frameworks co-exist in this area, offering subtly different takes. It is beyond the compass of this paper to mention them all. Van Benthem & Pacuit 2006 is a broad survey of temporal logics from mathematical, computational and philosophical traditions. Of special relevance here are epistemic versions of computational logics such as CTL and ATL (Fagin et al. 1995, van der Hoek & Wooldridge 2003), logics in the BDI tradition (Herzig & Longin 2004), and the STIT logic of deliberate temporal action in Belnap, Perloff & Xu 2001.<sup>20</sup>

## 7 Observation uncertainty and imperfect information games

---

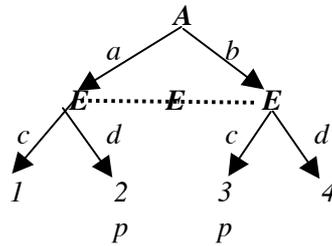
<sup>19</sup> For instance, we modeled a particular strategy by eliminating parts of game trees. But we could also have represented this by *expanding* the old game tree with new strategy relations.

<sup>20</sup> See van Benthem, Ghosh & Verbrugge, eds., to appear, for a view of the whole field.

We have now dealt with perfect information games, and uncertainty about how they are played. The next step are games with *imperfect information*, where players may not know exactly where they are in the game tree. The intuitive source for this kind of uncertainty is not procedural ignorance about the game one is in, but limitations on players' powers of *observation* or *memory*. There is a large literature on these games, and we just lightly review a few salient points from van Benthem 2001, 2011C.

*Example 24* Partial observation in games.

In this imperfect information game, the dotted line marks player *E*'s uncertainty about her position when her turn comes. She does not know the move played by player *A*:



**A local epistemic action language** Structures like this interpret a combined *dynamic-epistemic* language extending that of Section 5 on game trees. For instance, after *A* plays move *a* in the root, in both middle states, *E* knows ‘de dicto’ that playing *c* or *d* will give her *p* – as the disjunction  $\langle c \rangle p \vee \langle d \rangle p$  is true at both middle states:

$$K_E(\langle c \rangle p \vee \langle d \rangle p)$$

On the other hand, there is no specific move of which *E* knows ‘de re’ at this stage that it will guarantee a *p*-outcome – and this shows in the truth of the formula

$$\neg K_E \langle c \rangle p \wedge \neg K_E \langle d \rangle p$$

Such distinctions are typical when action and knowledge combine for agents.

**Strategies and knowledge** A striking new feature of our game is ‘non-determinacy’. *E*'s playing the opposite of player *A* is a strategy guaranteeing outcome *p* in the underlying perfect information game – but it is unusable now, since *E* cannot tell what *A* played. We want *uniform strategies* prescribing the same move at indistinguishable nodes. But then *E* has no strategy for forcing *p*, while *A* has no strategy forcing  $\neg p$ . Uniformity of strategies is guaranteed by special ‘knowledge programs’ in the

combined epistemic dynamic language over these games (cf. Fagin et al. 1995), whose test conditions for actions are propositions that an agent knows to be true or false.

**Styles of play and types of player** Again, there is dynamics lurking behind the statics of imperfect information games. Not just any sprinkling of uncertainty links over a perfect information game tree makes sense. In particular, one may want to assume that the only uncertainty arising comes from limited observation (as in card games), not from limited memory. The logic tells us the cash value of the earlier Perfect Recall:

*Fact 25* The axiom  $K_i[a]\varphi \rightarrow [a]K_i\varphi$  holds for player  $i$  for all propositions  $\varphi$

iff  $\mathcal{M}$  satisfies *Confluence*:  $\forall xyz: ((xR_a y \wedge y \sim_i z) \rightarrow \exists u: ((x \sim_i u \wedge uR_a z))$ .<sup>21</sup>

The converse  $[a]K_i\varphi \rightarrow K_i[a]\varphi$  holds for agents  $i$  satisfying No Miracles, whose epistemic uncertainty can only disappear by observing events they can distinguish.<sup>22</sup> The game of Example 24 has Perfect Recall, but No Miracles fails:  $E$  suddenly knows where she is after she played her move. Let us now make this dynamics explicit.

**Imperfect information games and dynamic-epistemic logic** Observation uncertainty is the typical province of dynamic-epistemic logics (Baltag, Moss & Solecki 1998, van Ditmarsch, van der Hoek & Kooi 2007, van Benthem 2011C). Here partly private events occur that may be observed differently by different agents, as encoded in ‘event models’  $E$  that form products with current epistemic models  $\mathcal{M}$ . We will not give any details of this ‘product update’ framework here, but merely mention one result that also points to a connection with the epistemic forest models of the preceding section. Which imperfect information games are playable by agents powered by *DEL*?

*Theorem 26* An extensive imperfect information game is isomorphic to an iterated

*DEL* update model  $Tree(\mathcal{M}, E)$  over some epistemic event model  $E$  iff

it satisfies, for all players, (a) Perfect Recall, (b) No Miracles, and  
(c) Bisimulation Invariance for the domains of moves.

For definitions of these notions and a proof of the result, see van Benthem 2001. Van Benthem, Gerbrandy, Hoshi & Pacuit 2009 prove the result for epistemic forest

---

<sup>21</sup> Similar analyses work for epistemic effects of memory bounds (van Benthem 2001).

<sup>22</sup> See Halpern & Vardi 1989 on these same issues in epistemic-temporal logic.

models.<sup>23</sup> The upshot of this technical line is as in our earlier discussion: we want to analyze the sources of different kinds of knowledge in games, and make them explicit.

**The total picture: epistemic forest models once more** Imperfect information games as logical models fit quite well into the epistemic forest models of the preceding section. Again they are a special case, but now virtually the opposite of an earlier one:

*Fact 27* Epistemic forest models for imperfect information games satisfy this condition: if  $h \sim h'$  (for any player  $i$ ), then  $h (\text{move} \cup \text{move}^v)^* h'$ .<sup>24</sup>

**Discussion: epistemic linear time** One may wonder whether the branching epistemic structure of epistemic forests is really needed here. Could not we trade off ‘forward’ uncertainty in a game tree with ‘sideways’ epistemic uncertainty between linear histories, as in the interpreted systems of Fagin, Halpern, Moses & Vardi 1995?

*Example 28* Trade-off.

Intuitively, the following two models seem close (Pacuit 2007. van Benthem 2009):



Indeed, in a branching temporal language, many assertions can be translated across. ■

In summary, we have seen that epistemic forests combine all sources of knowledge and ignorance in games: branching future, procedural information, and observation or memory. They support dynamic epistemic temporal languages that express elaborate scenarios of social action. But most of all, moving stepwise, we have shown how much fine-structure there is in the stages of coming to this most general framework.

**Yet more complex models of games** Epistemic forests are not the last word in the literature on logic and games. There are also more complex abstract *epistemic game models*  $\mathbf{M} = (W, \sim_i, V)$  (cf. Stalnaker 1999) whose worlds carry local information about all nodes in a game  $\mathbf{G}$ , plus strategy profiles specifying each player’s behavior

<sup>23</sup> For some interesting critical footnotes, see Dégrémont, Loewe & Witzel 2011.

<sup>24</sup> Again this can be defined in a hybrid version of our epistemic action language.

throughout the game. Players' information about structure and procedure is then encoded by uncertainty relations  $\sim_i$  between worlds. We will not discuss here just how this extends epistemic forests, or for what sorts of discourse this step upward is really needed – one instance are complex counterfactual assertions about strategies.

## 8 From knowledge to belief

Here is a natural test on the preceding styles of modeling knowledge. They also make sense for beliefs: as we have seen, the more common fuel driving games and social action. To show this, we will walk lightly along some earlier topics.

**Belief in tree models** Our first sort of knowledge in games was based on branching game trees. If we want to represent belief in the same models, there is a difference: expectations about the future seem largely about the future history, not single nodes. This can be modeled in the same style as for the temporal language of Section 5, using an earlier-mentioned device for dealing with belief. We add binary relations  $\leq_i$  of relative plausibility between histories, and then interpret a doxastic modality:

$$\mathbf{M}, h, s \models \langle B, i \rangle \phi \text{ iff } \mathbf{M}, h', s \models \phi \text{ for some history } h' \text{ coinciding with } h \text{ up to stage } s \text{ and most plausible for } i \text{ according to the given relation } \leq_i. \text{ }^{25}$$

Moving along a history, the plausibility relations encode belief changes for players like the ones we saw with Backward Induction in Section 3. This now shows in the logic, provided we also add an operator for conditional beliefs:

$$\mathbf{M}, h, s \models \langle B, i \rangle (\psi, \phi) \text{ iff } \mathbf{M}, h', s \models \phi \text{ for some history } h' \text{ coinciding with } h \text{ up to stage } s \text{ and most plausible for } i \text{ among all future histories satisfying } \psi.$$

We just state the next result to show how a temporal logic of belief works. Here the modal-temporal operator  $F_a$  says that the current branch starts with an  $a$  move:

*Fact 29* The following principles are valid for belief change:

$$\begin{aligned} \langle B, i \rangle F_a T &\rightarrow (F_a \langle B, i \rangle \phi \leftrightarrow (F_a T \ \& \ \langle B, i \rangle F_a \phi)) \\ F_a \langle B, i \rangle \phi &\leftrightarrow (F_a T \ \& \ \langle B, i \rangle (F_a T, F_a \phi)) \end{aligned} \text{ }^{26}$$

---

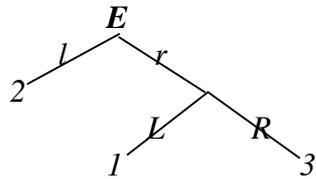
<sup>25</sup> If one wants to model significant beliefs of players about beliefs of others in this manner, then single trees will not suffice, and we need forests as before.

For similar observations in the setting of analyzing *AGM* theory, see Bonanno 2007, and for more extensive study of doxastic-temporal languages, see Dégrémont 2010.

***Belief among moves after all?*** In Section 3, Backward Induction *constructed* the plausibility ordering relations that were just taken as primitives in the above doxastic-temporal models. In the *BI* procedure as we analyzed it, plausibility relations matched strategies viewed as subrelations *best* of ‘best action’ inside the total *move* relation (Fact 5). But with such a local encoding of global plausibility, we can have a language for belief adding just a modality  $\langle \textit{best} \rangle$  to a propositional dynamic logic over game trees (Section 2). Then, for instance, we can extend the epistemic *PDL* strategies of Section 7 to doxastic ones involving belief conditions. Still, is this view reasonable? It becomes hard to define conditional beliefs, since we need gradations among non-best moves. But a more important objection is the uniformity imposed by the *BI*-algorithm:

*Example 30* Crossing expectations.

Recall the following picture from Example 1, but this time, with the numbers at the outcomes read, not as utility values, but as degrees of plausibility:



This pattern goes against the ‘node-compatible’ plausibility order for *BI* in Section 3, which makes one of the moves  $l, r$  more plausible than the other, while all their later outcomes follow this earlier decision. Yet it is easy to think of scenarios where the ‘crossing’ plausibility order depicted here is natural, and hence our general doxastic-temporal models seem the right level of generality. The special uniform plausibility relations produced by the *BI* algorithm induce a definable subclass satisfying a special axiom in the doxastic-temporal language, whose details we do not spell out here. ■

***From trees to epistemic-doxastic forests*** The same reasons that applied to knowledge also motivate a step from game trees to more complex epistemic doxastic forest

---

<sup>26</sup> To be complete, we would also have to formulate laws for new conditional beliefs.

models. First, agents may have beliefs rather than knowledge about the structure of the game. They may also have beliefs about other agents' different beliefs or knowledge. And finally, extending imperfect information games, they may have restricted powers of observation that generate beliefs in addition to knowledge about where they are in a game tree. For the resulting models and their temporal logics, see Dégrémont 2010.

**Update actions** As before, there are dynamic logics that describe effects on epistemic doxastic forest models of the key acts of observing moves and public announcements. In particular, public announcements now find a use in analyzing the earlier procedure of Forward Induction (cf. Example 2 and Sections 2, 3). When we observe a move  $e$ , then, taking  $e$  to be rational gives information about the active player  $i$ 's beliefs: they are such as to make  $e$  rational in beliefs. Or more succinctly, the update is for

! “move  $e$  is rational for  $i$  in beliefs”.

But in addition to this flow of ‘hard information’, our richer doxastic setting also supports new kinds of event that provide ‘soft information’ by changing a current plausibility order without eliminating worlds. In particular, Definition 10 introduced a transformation of ‘radical upgrade’  $\uparrow\varphi$  upgrading the proposition by putting the set of worlds satisfying  $\varphi$  on top in the plausibility ordering. A sophisticated mechanism of Priority Update, similar in scope to the dynamic-epistemic product update of Section 7, was defined in Baltag & Smets 2008. Van Benthem & Dégrémont 2008 prove an analogue to Theorem 27 showing which doxastic forests arise in this way. All this generalizes the earlier epistemic models with interesting new features.<sup>27</sup>

**Belief versus expectations** The only point we make here is a conceptual one. In our treatment of knowledge, but even more clearly with belief, we are really modeling two different notions. The first, inside game trees, is *expectation* about the future course of the game, the other, between trees, may be called *belief* proper (about the process, about other agents). Though the two notions are close, and both involve plausibility comparisons, the distinction seems useful and illuminating. Perhaps the usual broad category ‘belief’ is not fine-grained enough to understand strategic social behavior.

**A challenge: forward induction revisited** To make the preceding discussion a bit more concrete, consider Examples 1 and 2 once more, read in terms of plausibility. We start with a technical issue, that is at the same time conceptual. If the agent moved left by choosing  $a$ , we postulated an order for her on the second node where outcome 3 became less plausible than outcome 1. But if the agent moved right by choosing  $b$ , we assumed that 3 was more plausible than 1 for her. Thus, plausibility is not a uniform binary relation, as with Backward Induction. It may differ when viewed from different perspectives, making it a *node-dependent ternary* relation (van Benthem 2011A). This move is in line with semantics for conditional logic, and one can study its effects on valid laws of reasoning, which will now become less uniform than before.

But is this technical move to ternary plausibility coherent? Continuing with the conditional logic angle, it has been suggested that reasoning about behavior in games is held in place by counterfactuals. But then, should not the current plausibility order tell us already counterfactually what is most plausible at other nodes? “I chose move  $a$ , and now I say that, if I had chosen move  $b$ , the following would have happened... Where is the need for making the latter reasoning dependent on being at the node reached by playing  $b$ ?” There is a confusion here. If I have actually taken  $a$ , then I consider choosing  $b$  a *mistake*, and my counterfactual has an assumption that I would have gone there by mistake. But if I have actually taken  $b$ , the favored assumption is that I chose that move *intentionally*. In other words, the counterfactual is incomplete unless we add assumptions about how the move was played: intentionally, by mistake, and so on. And these assumptions generate different plausibility orderings.<sup>28</sup>

**Richer updates** The same issue is also technical. The proper way of observing moves in social scenarios may be by not taking them at face value, but in more intentional ways. Update actions are then more complex than public observations *!e*, such as “ $e$  was played on the basis of rationality in beliefs”, “ $e$  was played by mistake”, and so

---

<sup>27</sup> In particular, the limit and fixed-point theory of iterated soft updates, as used in our analysis of *BI*, is more delicate than that of repeated public announcements: cf. Baltag & Smets 2009.

<sup>28</sup> With Backward Induction, the usual assumption is that all off-path moves are mistakes, but the player will ‘do better from now on’, following the *BI* rule again.

on. Moreover, since these additional features may be hypotheses on our part, we may not want to use hard public announcements, but the earlier soft upgrade  $\uparrow$ .<sup>29 30</sup>

**Discussion: ‘keep it simple’** What is at stake here is again the issue of where to put what in a logical analysis of social action. I feel that it is always good policy to watch out for models growing heavier (van Benthem 2011B). Despite my sympathy for forests with procedural information and agent types, the above also suggests a local alternative, refocusing attention from global types and long-term behavior to individual moves, and local update responses to these.<sup>31</sup> But then, I have only made a start: what is the natural repertoire of such local events in the actual course of a game?

Modeling behavior in the presence of belief is a matter of conceptual analysis as much as logical technique. Once we have achieved the former, the latter will be forthcoming.

---

<sup>29</sup> Using  $\uparrow$  has the additional virtue that we can now make sense of any move, even those that are not rationalizable. A radical update for rationality in beliefs will put those worlds on top where the latter property holds, but when there are no such worlds, it will leave the plausibility order the same. What happens in that case is just a bare observation of the move. Of course, this minimal procedure does not solve the issue of how to solve conflicts in our interpretation of behavior, which may involve other updates in terms of *changing preferences* (Liu 2011).

<sup>30</sup> One obvious further issue is to which extent the stated update scenarios on forest models can be compressed downward to work on just game trees with plausibility orderings, the way things were with Backward Induction. Indeed one can give rough versions of Forward Induction where a plausibility order gets created directly on branches. For instance, start from a flat plausibility order, and consider successive nodes in the game tree, where possible moves partition the reachable outcomes. Upgrade all partition cells using a radical upgrade for the set of outcomes that majorize, for the active player, at least one outcome for an alter-native move. (What gets demoted out of the ‘belief range’ in this way are those outcomes that would be bad no matter what.) It is of interest to apply this simple procedure to the game of Example 2. We leave a precise statement of these simplified methods for later work.

<sup>31</sup> In evolutionary game theory, local response strategies like ‘Tit-for-Tat’, that refuse to think deeply about other players’ types, often perform best.

## 9 Conclusions and further directions

As we said at the start, this paper is a survey of some of the many logical models for games and social action, trying to see some connecting themes.<sup>32</sup> Our main point has been that this variety is not a drawback, but an asset. Instead of looking for huge worst-case models wearing winter clothes in summer, it seems better to let relevant scenarios dictate the things you want to capture, at the right level of lightness.

There is much more to the story of fine-structure and natural steps in game modeling than what we have been able to address in this paper, and more can be said also about combining existing frameworks (cf. van Benthem 2012 for further themes). In what follows, we mention a few points that deserve further study in our grand picture.

***Dynamics and pre-encoding*** It has been suggested that making the dynamics of play explicit, as we have done, is a natural complement to looking for small models (cf. Baltag, Smets & Zvesper 2009). One relocates part of the relevant information in the update process. How far does this go? Does it also work both ways? Can we think of epistemic-temporal ‘grand stage’ models (van Benthem & Pacuit 2006) as doing the reverse: simplify update to observing moves, but on a much larger static structure?

***Transforming different models*** There is a persistent intuition that the different kinds of knowledge that we have distinguished can be turned into each other through some sort of representation. For instance, we discussed turning ‘future ignorance’ based on branching in a game tree into uncertainty between histories in the style of epistemic logic? There exist some folklore results to this effect (cf. van Benthem 2009, 2011C) that also support some translations at the level of the above languages. But I am not aware of a systematic study clarifying just how far these changes in perspective go.

***Game change*** While the entanglement of action, knowledge, belief and preference in social reasoning is an exciting force behind this paper, the resulting notions may be very fragile. Changing a game a little bit may result in dramatic shifts in its strategic equilibria, and earlier best behavior may easily fall apart. While our logics provide tools for reasoning about many game changes, it is clear that we do not yet understand the true continuities and discontinuities in changing social scenarios.

---

<sup>32</sup> Our survey is by no means complete. For a bunch of very recent themes, cf. Turrini 2011.

**Long-term behavior** Many dynamic analyses of game-theoretic solution procedures involve limit behavior over time (cf. Dégrémont & Pacuit 2009, Pacuit & Roy 2011) that goes beyond single update steps, finding interesting stable models eventually. There is no systematic logical theory of all this yet, though Gierasimczuk 2010 makes some interesting connections with the formal learning theory of Kelly 1996.

**Groups** The epistemic foundations of game theory typically involve group notions like common knowledge or common belief. It is not at all trivial to generalize everything we have discussed to this setting. Finding natural recursion laws for common knowledge in *DEL* was a difficult task (van Benthem, van Eijck & Kooi 2006), and the same question for common belief under radical upgrade is still an open problem.

**Other kinds of update in games** We have discussed observing moves and receiving public information as driving forces in actual play. But there can be many others. Liu 2011 studies *preference changes* as another natural phenomenon, when players adjust their preferences in play: something that is well-documented in the cognitive science literature. This should be incorporated into the analysis that we have presented here.

We see this long list of things to do as a virtue of the ‘slow steps’ analysis in modeling for games that we have proposed. Taking the perspective of this paper draws together ideas from many logical frameworks, and it makes us see new issues all around.

## 10 References

- R. Aumann, 1995, ‘Backward Induction and Common Knowledge of Rationality’, *Games and Economic Behavior* 8, 6–19.
- A. Baltag, L. Moss & S. Solecki, 1998, ‘The Logic of Public Announcements, Common Knowledge and Private Suspicions’, *Proceedings TARK 1998*, 43–56, Morgan Kaufmann Publishers, Los Altos.
- A. Baltag & S. Smets, 2008, ‘A Qualitative Theory of Dynamic Interactive Belief Revision’, in G. Bonanno, W. van der Hoek, M. Wooldridge, eds., *Texts in Logic and Games Vol. 3*, Amsterdam University Press, 9–58.
- A. Baltag & S. Smets, 2009, ‘Group Belief Dynamics under Iterated Revision: Fixed Points and Cycles of Joint Upgrades’, *Proceedings TARK XII*, Stanford, 41–50.

- A. Baltag, S. Smets & J. Zvesper, 2009, 'Keep 'Hoping' for Rationality: a Solution to the Backward Induction Paradox', *Synthese* 169, 301–333.
- P. Battigalli & M. Siniscalchi, 2002, 'Strong Belief and Forward Induction Reasoning', *Journal of Economic Theory* 106:2, 356–391.
- N. Belnap, M. Perloff & M. Xu, 2001, *Facing the Future*, Oxford University Press, Oxford.
- J. van Benthem, 2000, *Logic in Games*, lecture notes, ILLC, University of Amsterdam. Downloadable version available, September 2011.
- J. van Benthem, 2001, 'Games in Dynamic Epistemic Logic', *Bulletin of Economic Research* 53:4, 219–248.
- J. van Benthem, 2002, 'Extensive Games as Process Models', *Journal of Logic, Language and Information* 11, 289–313.
- J. van Benthem, 2006, 'Rationalizations and Promises in Games', *Philosophical Trends*, 'Supplement 2006' on logic, Chinese Academy of Social Sciences, Beijing, 1–6.
- J. van Benthem, 2007A, Dynamic Logic of Belief Revision, *Journal of Applied Non-Classical Logics* 17:2, 129–155.
- J. van Benthem, 2007B, 'Rational Dynamics and Epistemic Logic in Games', *International Game Theory Review* 9:1, 2007, 13–45. (Erratum reprint, Volume 9:2, 377–409.)
- J. van Benthem, 2009, 'The Information in Intuitionistic Logic', *Synthese* 167:2, 251–270.
- J. van Benthem, 2010, 'In Praise of Strategies', In J. van Eijck and R. Verbrugge, eds., *Foundations of Social Software*, Studies in Logic, College Publications, London, 283–317.
- J. van Benthem, 2011A, 'Exploring a Theory of Play', invited lecture, in K. R. Apt, ed., *Proceedings TARK Groningen 2011*, ACM Digital Library, 12–16.
- J. van Benthem, 2011B, 'Keep it Simple', talk at a Workshop on *Recent Developments in Logical Dynamics*, Department of Philosophy, Peking University.
- J. van Benthem, 2011C, *Logical Dynamics of Information and Interaction*, Cambridge University Press.
- J. van Benthem, 2012, *Logic in Games*, ILLC, University of Amsterdam.

- J. van Benthem & C. Dégrémont, 2008, ‘Multi-Agent Belief Dynamics: bridging between dynamic and temporal doxastic logics’, *Proceedings LOFT Amsterdam*, 151–173.
- J. van Benthem, J. van Eijck & B. Kooi, 2006, ‘Logics of Communication and Change’, *Information and Computation* 204(11), 1620–1662.
- J. van Benthem, J. Gerbrandy, T. Hoshi and E. Pacuit, 2009, ‘Merging Frameworks for Interaction’, *Journal of Philosophical* 38:5, 491–526.
- J. van Benthem & A. Gheerbrant, 2010, ‘Game Solution, Epistemic Dynamics, and Fixed-Point Logics’, *Fundamenta Informaticae* 100, 19–41.
- J. van Benthem, S. Ghosh & R. Verbrugge, eds., to appear, *Strategies for Intelligent Interaction*, STRATMAS Project, <http://www.ai.rug.nl/~sujata/stratmas.html>
- J. van Benthem & E. Pacuit, 2006, ‘The Tree of Knowledge in Action: Towards a Common Perspective’, in G. Governatori, I. Hodkinson & Y. Venema, eds., *Proceedings AiML Melbourne 2006*, College Publications, London, 1–20.
- J. van Benthem, S. van Otterloo & O. Roy, 2006, ‘Preference Logic, Conditionals, and Solution Concepts in Games’, in H. Lagerlund, S. Lindström & R. Sliwinski, eds., *Modality Matters*, University of Uppsala, 61–76.
- C. Bicchieri, 1988, ‘Common Knowledge and Backward Induction: A Solution to the Paradox’, *Proceedings TARK 1988*, 381–393, Morgan Kaufman Publishers.
- P. Blackburn, M. de Rijke & Y. Venema, 2001, *Modal Logic*, Cambridge University Press.
- G. Bonanno, 2007, ‘Axiomatic Characterization of the AGM theory of Belief Revision in a Temporal Logic’, *Artificial Intelligence* 171, 144–160.
- A. Brandenburger, 2007, ‘Forward Induction’, Stern School of Business, NYU New York.
- C. Dégrémont, 2010, *The Temporal Mind: observations on logics of belief change in interactive systems*, Dissertation, ILLC, University of Amsterdam.
- C. Dégrémont, B. Löwe & A. Witzel, 2011, ‘The Synchronicity of Dynamic Epistemic Logic’, *Proceedings TARK 2011*, Groningen.
- C. Dégrémont & O. Roy, 2009, ‘Agreement Theorems in Dynamic Epistemic Logic’, In A. Heifetz, ed., *Proceedings TARK 2009*, Stanford, 91–98.
- H. van Ditmarsch, W. van der Hoek & B. Kooi, 2007, *Dynamic-Epistemic Logic*, Cambridge University Press, Cambridge.
- R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge (Mass.).

- N. Gierasimczuk, 2010, *Knowing One's Limits, Logical Analysis of Inductive Inference*, Dissertation, ILLC, University of Amsterdam.
- P. Girard, 2008, *Modal Logic for Belief and Preference Change*, Dissertation, Department of Philosophy, Stanford University.
- J. Halpern & M. Vardi, 1989, 'The Complexity of Reasoning about Knowledge and Time, I: lower bounds'. *Journal of Computer and System Sciences* 38,195–237.
- A. Herzig & D. Longin, 2004, 'C&L Intention Revisited', *Proceedings KR 2004*, AAAI-Press, 527–535.
- W. van der Hoek & M. Wooldridge, 2003, 'Cooperation, Knowledge, and Time. Alternating-time Temporal Epistemic Logic and its Applications', *Studia Logica* 75, 125–157.
- T. Hoshi, 2009, *Epistemic Dynamics and Protocol Information*, Ph.D. thesis, Department of Philosophy, Stanford University (ILLC-DS-2009-08).
- K. Kelly, 1996, *The Logic of Reliable Inquiry*, Oxford University Press, Oxford.
- F. Liu, 2011, *Reasoning about Preference Dynamics*, Synthese Library, Springer Science Publishers.
- E. Lorini & F. Moisan, 2011, 'An Epistemic Logic of Extensive Games', *Proceedings LAMAS 2011*, Electronic Notes in Theoretical Computer Science, Elsevier, 245–260.
- M. Osborne & A. Rubinstein, 1994, *A Course in Game Theory*, The MIT Press, Cambridge (Mass.).
- S. van Otterloo, W. van der Hoek, & M. Wooldridge, 2003, 'Knowledge as Strategic Ability', LCMAS workshop, Eindhoven, published in *ENTCS* 85:2, 2003.
- E. Pacuit, 2007, 'Some Comments on History Based Structures', *Journal of Applied Logic* 5:4, 613–624.
- E. Pacuit & O. Roy, 2011, 'A Dynamic Analysis of Interactive Rationality', in H. van Ditmarsch, J. Lang & S. Ju, eds., *Proceedings LORI III Guangzhou*, Springer Verlag, Lecture Notes in Computer Science.
- R. Parikh & R. Ramanujam, 2003, 'A Knowledge-Based Semantics of Messages', *Journal of Logic, Language and Information* 12, 453–467.
- A. Perea, 2011, 'Belief in the Opponents' Future Rationality', working paper, Epicenter, Department of Quantitative Economics, University of Maastricht.

- O. Roy, 2008, *Thinking Before Acting*, Dissertation Series DS-2008-03, ILLC, University of Amsterdam.
- R. Stalnaker, 1999, 'Extensive and Strategic Form: Games and Models for Games', *Research in Economics* 53, 293–291.
- C. Stirling, 1993, 'Modal and Temporal Logics', *Handbook of Logic in Computer Science*, Vol. 2, 477–563.
- P. Turrini, 2011, *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations*, SIKS Dissertation Series No. 2011-34, Department of Computer Science, University of Utrecht.