International Game Theory Review, Vol. 9, No. 1 (2007) 13–45

RATIONAL DYNAMICS AND EPISTEMIC LOGIC IN GAMES

JOHAN VAN BENTHEM

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam, Plantage Muidergracht 24
1018 TV AMSTERDAM, Netherlands
johan@science.uva.nl

Game-theoretic solution concepts describe sets of strategy profiles that are optimal for all players in some plausible sense. Such sets are often found by recursive algorithms like iterated removal of strictly dominated strategies in strategic games, or backward induction in extensive games. Standard logical analyses of solution sets use assumptions about players in fixed epistemic models for a given game, such as mutual knowledge of rationality. In this paper, we propose a different perspective, analyzing solution algorithms as processes of learning which change game models. Thus, strategic equilibrium gets linked to fixed-points of operations of repeated announcement of suitable epistemic statements. This dynamic stance provides a new look at the current interface of games, logic, and computation.

Keywords: Epistemic logic; dynamic logic; public announcement; rationality.

1. Reaching Equilibrium as an Epistemic Process

1.1. *Inductive solution algorithms for games*

Solving games often involves a stepwise algorithmic procedure. For instance, the well-known method of Backward Induction computes utility values at all nodes for players in finite extensive games in a bottom up manner. Strategic games in matrix form also support recursive algorithms. Our running example in this paper is

Example 1. Iterated removal of strictly dominated strategies (SD^ω).
Consider the following matrix, with this legend for pairs: (***A****-value,* ***E****-value)*.

	E *a*	*b*	*c*
A *d*	*2, 3*	*2, 2*	*1, 1*
e	*0, 2*	*4, 0*	*1, 0*
f	*0, 1*	*1, 4*	*2, 0*

Here is the method. First remove the dominated right-hand column: ***E***'s action c. After that, the bottom row for ***A***'s action f has become strictly dominated, and then, successively, ***E***'s action b, and then ***A***'s action e. The successive eliminations leave just the unique Nash game equilibrium (d, a) in the end.

In this example, SD^{ω} reaches a unique equilibrium profile. In general, it may stop at some larger solution zone of matrix entries where it performs no more eliminations.

1.2. *Solution methods and standard epistemic logic*

There is an extensive literature analyzing game-theoretic solution concepts in terms of epistemic logic, with major contributions by Aumann, Stalnaker, and many others. Just as a simple example, Binmore (1992) justifies the above steps in the SD^{ω} algorithm by means of *iterated mutual knowledge of rationality*:

> E.g., ***A*** can be sure that ***E*** will disregard the right-hand column, as ***A*** knows that ***E*** is rational. And ***E*** can later remove the second column since she knows that ***A*** knows her rationality, leading to discarding action c, and will therefore remove the bottom-most row. And so on.

The more complex the matrix, the more eliminations, and the greater the required depth of mutual knowledge. Technical characterization results behind this scenario show that the sets of profiles satisfying some given solution concept are just those that would occur in epistemic models for some suitable type of rationality assertion, involving mutual knowledge and beliefs of players. De Bruin (2004) has a survey of the mathematics of twenty years of such 'game logic' results, and their philosophical significance. In this paper we propose a new tack on these well-studied phenomena, emphasizing the dynamic nature of the algorithm. This reflects a change in epistemic logic since the 1980s.

1.3. *The dynamic turn in epistemic logic: Information update*

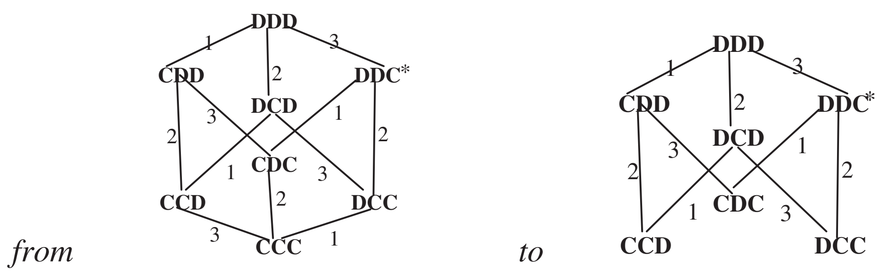
Standard epistemic logic describes what agents know, or don't, at worlds in some fixed situation. But normally, knowledge is the result of *actions* of observation, learning, or communication. In modern epistemic logics, such actions have become first-class citizens in system design. Van Benthem (1996) is a general investigation of this 'Dynamic Turn', which also shows in belief revision theories in AI, logics of interaction in computer science, and 'dynamic semantics' in linguistics. In this paper, solving a well-known knowledge puzzle will be our running illustration:

Example 2. 'Muddy Children' (Fagin *et al.*, 1995).
After playing outside, two of three children have mud on their foreheads. They can only see the others, so they do not know their own status. Now their Father says: "At least one of you is dirty". He then asks: "Does anyone know if he is dirty?" Children answer truthfully. As questions and answers repeat, what happens?

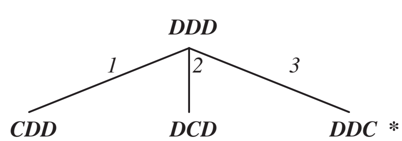
> Nobody knows in the first round. But in the next round, each muddy child can reason like this: "If I were clean, the one dirty child I see would have seen only clean children, and so she would have known that she was dirty at once. But she did not. So I must be dirty, too!"

In the initial epistemic model for this situation, eight possible worlds assign $\boldsymbol{D}$ or $\boldsymbol{C}$ to each child. A child knows about the others' faces, but not about her own, as reflected in the accessibility lines in the diagrams below, encoding agents' uncertainty. Now, the successive assertions made in the scenario *update* this information.

Example 2. (*continued*) Updates for the muddy children.
Updates start with the Father's public announcement that at least one child is dirty. This simple communicative action merely eliminates those worlds from the initial model where the stated proposition is false. I.e., $\boldsymbol{CCC}$ disappears:



When no one knows his status, the bottom worlds disappear:



The final update is to $\boldsymbol{DDC}*$.

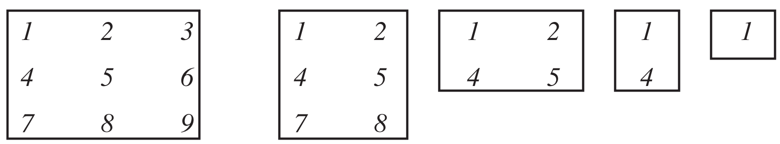
In this sequence of four epistemic models, domains decrease in size: 8, 7, 4, 1. With k muddy children, k rounds of stating the same simultaneous ignorance assertion "I do not know my status" by everyone yield common knowledge about which children are dirty. A few more assertions by those who now know achieve common knowledge of the complete actual distribution of the $\boldsymbol{D}$ and $\boldsymbol{C}$ for the whole group.

This solution process is driven by repeated announcement of the same assertion of ignorance, though its effect on the model is different every time it gets repeated. We will analyze this epistemic procedure in technical detail below. Clearly, there are much more sophisticated epistemic actions than merely announcing something in public — but this simple case will do for the rest of this paper.

1.4. *Solution methods as epistemic procedures*

There is an obvious analogy between our two examples so far. One might think of game-theoretic solution algorithms as mere tools to compute a Nash equilibrium, or some larger solution zone. But SD^{ω} and its ilk are also interesting processes in their own right, whose successive steps are epistemic actions changing game models to smaller ones. Initially, all options are still in, but gradually, the model changes to a smaller one, where players have more knowledge about possible rational outcomes.

Example 1. (*continued*) Updates for SD^{ω} rounds.
Here is the sequence of successive updates for the rounds of the algorithm:



Here each box may be viewed as an epistemic model. Each step increases players' knowledge, until some equilibrium sets in where they 'know the best they can'.

In Sec. 3, we shall see *which assertion* drives these elimination steps. Analyzing the algorithm in this detailed manner involves both epistemic logic and dynamic logic. In particular, in modern dynamic-epistemic logics (cf. Sec. 2), basic actions eliminating worlds from a model correspond to public announcements of some fact.

1.5. *Exploring the analogy*

This simple analogy between game solution algorithms and epistemic communication procedures is the main idea of this paper. It has surprising repercussions worth pursuing, even though it is not a panacea for all problems of rational action. Section 2 explains the machinery of dynamic-epistemic logic, including the intriguing behaviour of repeated assertions. Section 3 has the specific epistemic game models with preference structure that we will work with, and we explore their logic. In Sec. 4, we define two major options for 'rationality' of players, and find a complete description of the finite models of interest. Then in Sec. 5, we analyze SD^{ω} as repeated assertion of 'weak rationality', and also show how announcement procedures suggest a variant algorithm driven by 'strong rationality' which matches Rationalizability. Section 6 analyzes the resulting framework for game analysis in arbitrary models, and it proves that the solution zones for repeated announcement are definable in epistemic fixed-point logic. This links game-theoretic equilibrium theory with current fixed-point logics of computation. Finally, Sec. 7 points out further relevant features of the Muddy Children puzzle scenario, considering also epistemic procedures that revise players' beliefs. Section 8 generalizes our dynamic-epistemic analysis to extensive games and Backward Induction.

2. Dynamic Epistemic Logic

2.1. *Standard epistemic logic in a nutshell*

The language of standard epistemic logic has a propositional base with added modal operators $K_i\phi$ ('*i knows that* ϕ') and $C_G\phi$ ('ϕ is *common knowledge* in group G'):

$$p|\neg\phi|\,\phi\vee\psi|\,K_i\phi|\,C_G\phi$$

We also write $\langle j\rangle\phi$ for the dual modality of truth in at least one accessible world. This paper uses standard multi-$S5$ models $\boldsymbol{M}$ whose accessibilities are equivalence relations for agents. We write $(\boldsymbol{M}, s)$ for models with a current world s, suppressing brackets when convenient. Also, all models are *finite*, unless otherwise specified, though few of our results really depend on this.

The key semantic clauses are as follows:

$$\begin{aligned}\boldsymbol{M}, s \models K_i\phi \quad & \text{iff for all } t \text{ with } s\sim_i t, \boldsymbol{M} : t\models\phi\\ \boldsymbol{M}, s \models C_G\phi \quad & \text{iff for all } t \text{ that are reachable from } s \text{ by some}\\ & \qquad \text{finite sequence of } \sim_i \text{ steps (any } i\in G) : \boldsymbol{M}, t\models\phi\end{aligned}$$

A useful further technical device is 'relativized common knowledge' $C_G(\phi,\psi)$ (Kooi & van Benthem, 2004): ψ holds after every finite sequence of accessibility steps for agents going through ϕ-worlds only. This notion goes beyond the basic language.

Next, a basic model-theoretic notion states when two epistemic models represent the same information from the viewpoint of our standard epistemic language.

Definition 1. Epistemic bisimulation.
A *bisimulation* between epistemic models $\boldsymbol{M}$, $\boldsymbol{N}$ is a binary relation $\equiv$ between states m, n in $\boldsymbol{M}$, $\boldsymbol{N}$ such that, whenever $m\equiv n$, then (a) m, n satisfy the same proposition letters, (b1) if mRm', then there exists a world n' with nRn' and $m'\equiv n'$, (b2) the same 'zigzag clause' holds in the opposite direction.

Every model $(\boldsymbol{M}, s)$ has a smallest bisimilar model $(\boldsymbol{N}, s)$, its 'bisimulation contraction'. The latter is the simplest representation of the epistemic information in $(\boldsymbol{M}, s)$. Next, the following connection with our language holds.

Proposition 1. *Invariance for bisimulation.*
For every bisimulation E between two models $\boldsymbol{M}, \boldsymbol{N}$ with sEt, the worlds s, t satisfy the same formulas in the epistemic language with common knowledge.

Here is a converse to Proposition 1.

Theorem 1. *Epistemic definability of models.*
Each finite model $(\boldsymbol{M}, s)$ has an epistemic formula $\delta(\boldsymbol{M}, s)$ (with common knowledge) such that the following are equivalent for all models $\boldsymbol{N}$, t,

(a) $\boldsymbol{N}, t\models\delta(\boldsymbol{M}, s)$,
(b) *$\boldsymbol{N}, t$ has a bisimulation $\equiv$ with $\boldsymbol{M}, s$ such that $s\equiv t$.*

For a fast proof, cf. van Benthem (2002b). Thus there is a strongest epistemic assertion one can make about states in a current model. In particular, each world in a bisimulation contraction has a unique epistemic definition inside that model. For complete axiomatizations of epistemic validities, cf. Meijer & van der Hoek (1995).

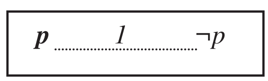
Remark 1. Distributed group knowledge.
Sometimes, we also need *distributed knowledge* in a group G:

$$\boldsymbol{M}, s \models D_G\phi \quad \text{iff for all } t \text{ with } s \cap_{i\in G} \sim_i t : \boldsymbol{M}, t \models \phi^1$$

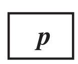
2.2. *Public announcement and model change by world elimination*

Now we extend this language to also describe events where information flows. Epistemic models change each time communication takes place. Such changes are crucial in 'dynamified' epistemic logic (van Benthem, 2002b). The simplest case is elimination of worlds from a model by public announcement of some proposition P.

Example 3. Questions and answers.
Here is an example. Some fact p is the case, agent 1 does not know this, while 2 does. Here is a standard epistemic model where this happens:



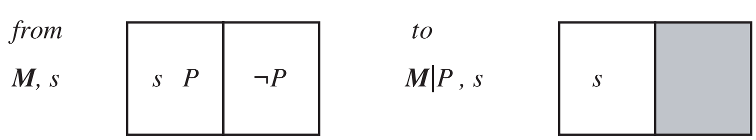
In the actual world $\boldsymbol{p}$, 1 does not know if p, but she knows that 2 knows. Thus, 1 can ask a question "p?". A truthful answer "*Yes*" by 2 then updates this model to



where p has become common knowledge in the group $\{1, 2\}$.

The general principle behind this simple epistemic model change is as follows.

Definition 2. Eliminative update for public announcement.
Let proposition P be true in the actual world of some current model $\boldsymbol{M}$. Truthful public announcement of P then removes all worlds from $\boldsymbol{M}$ where P does not hold, to obtain a new model $\boldsymbol{M}|P, s)$ whose domain is restricted to $\{t \in \boldsymbol{M} | \boldsymbol{M}, t \models P\}$:



[1]Bisimulation invariance fails for distributed knowledge (van Benthem, 1996).

State elimination is the simplest update action, with shrinking sets representing growing knowledge about the actual world. Atomic facts retain their truth in this process. But eliminative update may change the truth value of complex epistemic assertions ϕ at worlds, as we re-evaluate modalities in new smaller models. E.g., with the Muddy Children of Example 2, true statements about ignorance became false as worlds dropped out. Thus, in the end, common knowledge was achieved.

2.3. *Public announcement logic*

Update leads to in a *dynamic epistemic logic* using ideas from dynamic logic of programs to form mixed assertions allowing explicit reference to epistemic actions:

Definition 3. Logic of public announcement.
To all the formation rules of standard epistemic logic, we add a dynamic modality

$$[P!]\phi \quad \text{after truthful public announcement of } P, \text{ formula } \phi \text{ holds}$$

The truth condition is that $\boldsymbol{M}, s \models [P!]\phi$ iff, if $\boldsymbol{M}, s \models P$, then $\boldsymbol{M}|P, s \models \phi$.

This language can say things like $[A!]K_jB$: after truthful public announcement of A, agent j knows that B, or $[A!]C_GA$: after its announcement, A has become common knowledge in the group of agents G. Public announcement logic can be axiomatized completely. Typical valid principles describe interchanges of update actions and knowledge, relating so-called 'postconditions' of actions to their 'preconditions'.

Theorem 2. Completeness of public announcement logic.
Public announcement logic is axiomatized completely by (a) *all valid laws of standard epistemic logic*, (b) *the following five equivalences*:

$$[P!]q \leftrightarrow P \rightarrow q \quad \textit{for atomic facts } q$$

$$[P!]\neg\phi \leftrightarrow P \rightarrow \neg[P!]\phi$$

$$[P!]\phi \wedge \psi \leftrightarrow [P!]\phi \wedge [P!]\psi$$

$$[P!]K_i\phi \leftrightarrow P \rightarrow K_i[P!]\phi$$

$$[P!]C_G(\phi, \psi) \leftrightarrow C_G(P \wedge [P!]\phi, [P!]\psi)$$

We will not use this formal system in this paper, but it does provide the means of formalizing much of what we propose. (Cf. van Benthem, van Eijck & Kooi (2004) for a completeness proof, and applications.) But we also need an extension of its syntax.

2.4. *Program structure and iterated announcement limits*

Communication involves not just single public announcements. There are also sequential operations of *composition*, *guarded choice* and, especially, *iteration,* witness our Muddy Children story. Our main interest in this paper are public statements pushed to the limit. Consider any statement ϕ in our epistemic language. For any model $\boldsymbol{M}$ we can keep announcing ϕ, retaining just those worlds where ϕ holds. This yields an sequence of nested decreasing sets, which clearly stops in finite models. But also in infinite models, we can take the sequence across limit ordinals by taking intersections of all stages so far. Either way, the process must reach a *fixed-point*, i.e., a submodel where taking just the worlds satisfying ϕ no longer changes the model.

Definition 4. Announcement limits in a model.
For any model $\boldsymbol{M}$ and formula ϕ, the *announcement limit* $\#(\phi, \boldsymbol{M})$ is the first submodel in the repeated announcement sequence where announcing ϕ has no further effect. If $\#(\phi, \boldsymbol{M})$ is non-empty, ϕ has become common knowledge in this final model. We call such statements *self-fulfilling* in $\boldsymbol{M}$, all others are *self-refuting*.

The rationality assertions for games in Sec. 4 are self-fulfilling. The joint ignorance statement of the muddy children was self-refuting: inducing common knowledge of its negation. Both types of announcement limit are of interest.[2]

2.5. *Maximal group communication*

If muddy children tell each other what they see, common knowledge of the true world is reached at once. We now describe what a group can achieve by maximal public announcement. Agents in a model $(\boldsymbol{M}, s)$ can tell each other things they know, thereby cutting down the model to smaller sizes, until nothing changes.

Theorem 3. *Each model* $(\boldsymbol{M}, s)$ *has a unique minimal submodel reachable by maximal communication of known propositions among all agents. Up to bisimulation, its domain is the set* $COM(\boldsymbol{M}, s) = \{t | s \cap_{i \in G} \sim_i t\}$.

Proof. As this result nicely demonstrates our later conversation scenarios, we give a proof, essentially taken from van Benthem (2002). First, suppose agents reach a submodel $(\boldsymbol{N}, s)$ where further announcements of what they know have no effect. Without loss of generality, let $(\boldsymbol{N}, s)$ be contracted modulo bisimulation. Then Theorem 2 applies, and each world, and each subset, has an explicit epistemic definition. Applying this to the sets of worlds $\{t | s \sim_i t\}$ whose defining proposition is known to agent i, the agent could state this. But since this does not change the model, the whole domain is already contained in this set. Thus, $(\boldsymbol{N}, s)$ is contained in $COM(\boldsymbol{M}, s)$. Conversely, all of $COM(\boldsymbol{M}, s)$ survives each episode of public update, as agents only make statements true in all their accessible worlds.[3] □

[2] There is a role for an *actual world* s in our models. True announcements state propositions ϕ true at s in $\boldsymbol{M}$. If ϕ is self-fulfilling, its iterated announcement could have started at any world s in $\#(\phi, \boldsymbol{M})$. As s stays in at each stage, ϕ would have been true all the time. The limit consists of all worlds that could have been actual in this way. With this said, we suppress the s.

[3] Agents can reach $COM(\boldsymbol{M}, s)$ by speaking just once. We interleave bisimulation contractions with update, making sure all subsets are epistemically definable. First, agent 1 states all he knows by the disjunction $\vee\delta_t$ for all worlds t he considers indistinguishable from the actual one. This cuts the model down to the set $\{t | s \sim_1 t\}$. Next, it is 2's turn. But, the first update may have removed worlds that distinguished between otherwise similar worlds. So, we first take a bisimulation contraction, and then let 2 say the strongest thing he knows, cutting $\{t | s \sim_1 t\}$ down to those worlds that are also $\sim_2$-accessible from the actual one. Repeating this leads to $COM(\boldsymbol{M}, s)$.

$COM(\boldsymbol{M}, s)$ is the right set of worlds for evaluating distributed group knowledge-even though evaluating inside it does not quite yield the earlier notion $D_G\phi$.[4]

2.6. *Other epistemic actions*

Public announcement is the simplest form of communication. More sophisticated dynamic epistemic logics in the above style describe partial observation, hiding, misleading, and cheating, but they go beyond the simple scenarios of this paper (Baltag, Moss & Solecki (1998); van Ditmarsch, van der Hoek & Kooi (2005)).

3. Epistemic Logic of Strategic Game Forms

A dynamic epistemic analysis of game solution as model change needs static epistemic game models serving as group information states. In this paper, we choose a very simple version — to keep the general proposal as simple as possible, and make the dynamics itself the key feature. But all we say would also work for more sophisticated epistemic models of games found in the literature.

3.1. *Epistemic game models*

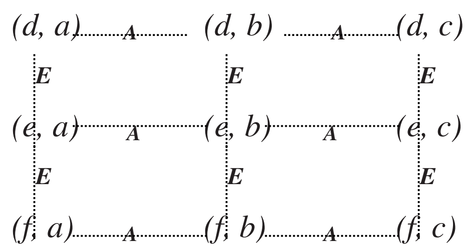
Consider a strategic game $\boldsymbol{G} = (I, \{A_j | j \in I\}, \{P_j | j \in I\})$ with a set of players I, and sets of actions A_j for each player $j \in I$. We shall mainly discuss *finite two-player games* — even though most results generalize. A tuple of actions for each player is a *strategy profile*, determining a unique outcome — and each player has his own *preference relation* P_j among these possible outcomes of the game. In this paper, we work with a minimal epistemic super-structure over such games:

Definition 5. The *full model over* $\boldsymbol{G}$ is a multi-$S5$ epistemic model $\boldsymbol{M}(\boldsymbol{G})$ whose worlds are all strategy profiles, and whose epistemic accessibility $\sim_j$ for player j is the equivalence relation of agreement of profiles in the j'th co-ordinate.

This stipulation means that players know their own action, but not that of the others. Thus, models describe the moment of decision for players having all relevant evidence. To model a more genuine process of deliberation, richer models would be needed — allowing for players' ignorance about other features as well. Such models exist in the literature, but we will stick with this simplest scenario. Likewise, we ignore all issues having to do with probabilistic combinations of actions. On our minimalist view, we can read a game matrix directly as an epistemic model.

[4]More delicate planning problems include announcing facts publicly between some agents while leaving others in the dark (cf. the 'Moscow Puzzle' in van Ditmarsch 2002).

Example 4. Matrix game models.
The model for the matrix game in Example 1 looks as follows:



Here ***E***'s uncertainty relation $\sim_{\boldsymbol{E}}$ runs along columns, because ***E*** knows his own action, but not that of ***A***. The uncertainty relation of ***A*** runs among the rows.

Solution algorithms like SD^{ω} may change such full game models to smaller ones:

Definition 6. A *general game model* ***M*** is any submodel of a full game model.

Omitting certain strategy profiles represents common knowledge between players of constraints on the global decision situation. This gives full logically generality, because of the following result (van Benthem (1996)):

Theorem 4. *Every multi-*$S5$ *model has a bisimulation with a general game model.*

Corollary 1. *The complete logic of general game models is just multi-*$S5$*.*

3.2. *Epistemic logic of game models*

Our game models support any type of epistemic statement in the languages of Sec. 2.1. Some examples will follow in Sec. 3.3. Even though this paper is not about complete logics, we mention some interesting validities in full game models. First, the matrix grid pattern validates the modal *Confluence Axiom* $K_AK_E\phi \to K_EK_A\phi$. For, with two players, any world can reach any other by composing the relations $\sim_A$ and $\sim_E$: so both K_AK_E and K_EK_A express universal accessibility. Next, the finiteness of our models implies upward well-foundedness of the two-step relation $\sim_A; \sim_E$, which has only ascending sequences of finite length. This validates a *Grzegorczyk Axiom* for the modality K_AK_E (cf. Blackburn, de Rijke & Venema (2000)). Such structural principles are crucial to modal reasoning about players' epistemic situations.[5]

[5] Actually, as to computational complexity, the logic of arbitrary full game models (finite or infinite) becomes *undecidable*, once we add common knowledge or a universal modality. The reason is that one can encode Tiling Problems on these grids. But the case for finite models only seems open.

3.3. *Best response and Nash equilibria*

To talk about solutions and equilibria we need further structure in game models, in particular, some atomic assertions that reflect the preferences underneath.

Definition 7. Expanded game language.
In an epistemic model $\boldsymbol{M}$ for a game, full or general, we say that player j performs action $\omega(j)$ in world ω, while $\omega(j/a)$ is the strategy profile ω with $\omega(j)$ replaced by action a. The *best response proposition* B_j for j says that j's utility cannot improve by changing her action in ω — keeping the others' actions fixed:

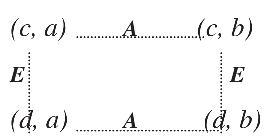
$$\boldsymbol{M}, \omega \models B_j \qquad \text{iff} \quad \mathop{\&}_{\{a \in Aj \mid a \neq \omega(j)\}} \; \omega(j/a) \leq_j \omega$$

Nash Equilibrium is expressed by the conjunction $NE = \mathop{\&} B_j$ of all B_j-assertions.

Stated in this way, best response is an *absolute* property referring to all actions in the original given game $\boldsymbol{G}$ — whether these occur in the model $\boldsymbol{M}$ or not. Thus, B_j is an atomic proposition letter, which keeps its value when models change.

Example 5. Expanded game models.
Well-known games provide simple examples of the epistemic models of interest here. Consider Battle of the Sexes with its two Nash equilibria. The abbreviated diagram to the right has best-response atomic propositions at worlds where they are true:

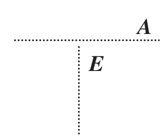
	E	*a*	*b*
A	*c*	2, 1	0, 0
	d	0, 0	1, 2



B_A, B_E	–
–	B_A, B_E

Next, our running Example 1 yields a full epistemic game model with 9 worlds:

	a	*b*	*c*
d	B_A, B_E	–	–
e	B_E	B_A	–
f	–	B_E	B_A



As for the distribution of the B_j-atoms, by the above definition, every column in a full game model must have at least one occurrence of B_A, and every row one of B_E.

Inside these models, more complex epistemic assertions can also be evaluated.

Example 5. (*continued*) Evaluating epistemic game statements.
(a) The formula $\langle \boldsymbol{E} \rangle B_E \wedge \langle \boldsymbol{A} \rangle B_A$ says that everyone thinks his current action might be best for him. In the 9-world model of our running example, this is true in exactly

the six worlds in the a, b columns. (b) The same model also high-lights an important epistemic distinction. B_j expresses that j's current action is *in fact* a best response at ω. But j need not *know* that, as she need not know what the other player is doing. Indeed, the statement $K_E B_E$ is false throughout the above model, even though B_E is true at three worlds.

A fortiori, then, common knowledge of rationality in its most obvious sense is often false throughout the full model of a game, even one with a unique Nash equilibrium.

With this enriched language, the logic of game models becomes more interesting.

Example 6. Valid game laws involving best response.
The following principle holds in all full game models: $\langle \boldsymbol{E} \rangle B_A \wedge \langle \boldsymbol{A} \rangle B_E$. It expresses the final observation of Example 5. We will see further valid principles later on.

But there are also other languages for game models. In particular, the above word 'best' is context-dependent. A natural *relative* version of best response in a general game model $\boldsymbol{M}$ looks only at the strategy profiles available *inside* $\boldsymbol{M}$. After all, in that model players know these are the only actions that will occur.

Definition 8. Relative best response.
The *relative best response* proposition B_j^* in a general game model $\boldsymbol{M}$ is true at only those strategy profiles where j's action is a best response to that of her opponent when the comparison set is all alternative strategy profiles in $\boldsymbol{M}$.

With B_j^*, best profiles for j may change as the model changes. E.g., in a one-world model for a game, the single profile is relatively best for all players, though it may be absolutely best for none. The relative version has independent interest:[6]

Remark 3. Relative best response and implicit knowledge.
Relative best response may be interpreted in epistemic terms. With two players, it says *the other player knows* that j's current action is at most as good for j as j's action at ω! More generally, B_j^* says that the proposition "j's current action is at most as good for j as j's action at ω" is *distributed knowledge* at ω for the rest of the group $G - \{j\}$. Intuitively, the others might learn this fact by pooling their information. This observation is used in van Benthem, van Otterloo & Roy (2005) for defining Nash equilibrium in an extended epistemic preference logic.

Assertions B_j^* return in our analysis of epistemic assertions driving SD^ω in Sec. 4. Finally, the connection: absolute-best implies relative-best, but not vice versa.

[6]The same distinction absolute versus relative best response also returns as a systematic choice point in the lattice-theoretic analysis of game transformations in Apt 2005.

Example 7. All models have relative best positions.
To see the difference between the two notions, compare the two models

1, 1 (B_A)	*0, 2* (B_E)
0, 2 (B_E)	*1, 1* (B_A)

1, 1 (B_A, B^*_E)	
0, 2 (B_E)	*1, 1* (B_A)

[7]

4. Rationality Assertions

Rationality plays one's best response given what one knows or believes. But our models support distinctions here, such as absolute versus relative best. Moreover, we found that even if players in fact play their best action, they need not *know* that they are doing so. So, if rationality is to be a self-reflective property, what *can* they know? This issue is also important in our epistemic conversation scenarios for game solution (Secs. 2, 5). Normally, we let players only say things which they know to be true.

4.1. *Weak rationality*

Players may not know that their action is best — but they can know that *there is no alternative action which they* know *to be better*. In short, 'they are no fools'.

Definition 9. Weak Rationality.
Weak Rationality at world ω in a model $\boldsymbol{M}$ is the assertion that, for each available alternative action, j thinks the current one may be at least as good

$$WR_j \quad \&_{a \neq \omega(j)} \langle j \rangle \text{ `}j\text{'s current action is at least as good for } j \text{ as } a\text{'}$$

Here the index set for the conjunction runs over just the worlds in the current model, as with relative best B^*_j in Sec. 3.3.[8]

The Weak Rationality assertion WR_j was defined to fail exactly at those rows or columns in a two-player general game model that are strictly dominated for j. E.g., unpacking the quantifiers in Definition 9, in our running Example 1, WR_E says for a column x that for each other column y, there is at least one row where $\boldsymbol{E}$'s value in x is at least as good as that in y. Evidently, such columns always exist.

Theorem 5. *Every finite general game model has worlds with the statement WR_j true for all players j.*

[7]The original version of this paper (van Benthem 2002C) has a richer language for game models with preference modalities, nominals for worlds, a universal modality, and distributed group knowledge. These modal gadgets yield formulas like $K_A \downarrow \langle \boldsymbol{E} \rangle act_E \geq_E \downarrow act_E$ which formalize rationality principles. Van Benthem, van Otterloo & Roy (2005) explore such preference logics for games. In what follows, however, we deal with our models in a more informal manner.
[8]An alternative version would let the index set run over all strategy profiles in the whole initial game — as happened with absolute best assertions B_j. It can be dealt with similarly.

Proof. For convenience, look at games with just two players. We show something stronger, viz. that the model has *WR-loops* of the form

$$s_1 \sim_{\boldsymbol{A}} s_2 \sim_{\boldsymbol{E}} \ldots \sim_{\boldsymbol{A}} s_n \sim_{\boldsymbol{E}} s_1 \qquad \text{with } s_1 |= B^*_{\boldsymbol{E}},\ s_2 |= B^*_{\boldsymbol{A}},\ s_3 |= B^*_{\boldsymbol{E}}, \ldots$$

By way of illustration, a Nash equilibrium by itself is a one-world *WR*-loop. First, taking maxima on the available positions (*column, row*) in the full game matrix, we see that the following two statements must hold everywhere (cf. Sec. 3):

$$\langle \boldsymbol{E} \rangle B^*_A \qquad \langle \boldsymbol{A} \rangle B^*_E$$

The first says that, given a world with some action for $\boldsymbol{E}$, there must be a world in the model with that same action for $\boldsymbol{E}$ where $\boldsymbol{A}$'s utility is highest. (This need not hold with the above absolute B_A, as its 'witness world' may have been left out.) Repeating this, there is a never-ending sequence of worlds $B^*_{\boldsymbol{E}} \sim_{\boldsymbol{E}} B^*_{\boldsymbol{A}} \sim_{\boldsymbol{A}} B^*_{\boldsymbol{E}} \sim_{\boldsymbol{E}} B^*_{\boldsymbol{A}} \ldots$ which must loop since the model is finite. Thus, some world in the sequence with, say, $B^*_{\boldsymbol{A}}$ must be $\sim_{\boldsymbol{A}}$ — connected to some earlier world w. Now, either w has $B^*_{\boldsymbol{E}}$, or w has a successor with $B^*_{\boldsymbol{E}}$ via $\sim_{\boldsymbol{A}}$ in the sequence. The former case reduces to the latter by the transitivity of $\sim_{\boldsymbol{A}}$. But then, looking backwards along such a loop, and using the symmetry of the accessibility relations, we have a *WR*-loop as defined above. Its worlds evidently validate Weak Rationality for both players: $\langle \boldsymbol{E} \rangle B^*_E \wedge \langle \boldsymbol{A} \rangle B^*_A$. □

Proposition 2. *Weak Rationality is epistemically introspective.*

Proof. By Definition 9 and accessibility in epistemic game models, if WR_j holds at some world ω in a model, it also holds at all worlds that j cannot distinguish from ω. Hence, the principle $WR_j \rightarrow K_j\, WR_j$ is valid on general game models. □

Thus, $WR_j \rightarrow K_j\, WR_j$ is a logical law of game models with best response and rationality. This makes Weak Rationality a suitable assertion for public announcement, ruling out worlds on strictly dominated rows or columns every time when uttered.

4.2. *Strong rationality*

Weak Rationality is a logical conjunction of epistemic possibility operators: *&* $\langle j \rangle$. A stronger form of rationality assertion would invert this order, expressing that players think that *their actual action may be best*. In a slogan, instead of merely 'being no fool', they can now reasonably 'hope they are being clever'.

Definition 10. Strong Rationality.
Strong rationality for j at a world ω in a model $\boldsymbol{M}$ is the assertion that j thinks that her current action may be at least as good as all others:

$$SR_j \qquad \langle j \rangle\ \&_{a \neq \omega(j)} \text{`}j\text{'s current action is at least as good for } j \text{ as } a\text{'}$$

This time we use the absolute index set running over all action profiles in the game. This means that the assertion can be written equivalently as the modal formula $\langle j \rangle B_j$. Strong Rationality for the whole group of players is the conjunction $\&_j SR_j$.

By the $S5$-law $\langle j \rangle \phi \to K_j \langle j \rangle \phi$, SR_j is something that players j will know if true. Thus, it behaves like WR_j. Moreover, we have this comparison:

Proposition 3. *SR_j implies WR_j, but not vice versa.*

Proof. Consider the following game model, with B-atoms indicated:

	E	*a*	*b*	*c*
A	*d*	*1, 2*	*1, 0*	*1, 1*
	e	*0, 0*	*0, 2*	*2, 1*

B_E, B_A	B_A	–
–	B_E	B_A

No column or row dominates any other, and WR_j holds throughout for both players. But SR_E holds only in the two left-most columns. For it rejects actions which are never best, even though there need not be one alternative better over-all. □

One advantage of SR_j over WR_j is the absoluteness of the proposition letters B_j in its definition. Once these are assigned, we never need to go back to numerical utility values for computing further stages of iterated announcement. In our later dynamic analysis in Sections 5 and 6, this feature underlies the semantic *monotonicity* of the set transformation defined by SR, a point made independently in Apt 2005. Strong Rationality has a straightforward game-theoretic meaning:

> *The current action of the player is a best response against at least one possible action of the opponent.*

This is precisely the assertion behind 'rationalizability' views of game solution, due to Bernheim and Pearce (de Bruin, 2004; Apt, 2005), where one discards actions for which a better response exists under all circumstances. We return to this later.

Strong Rationality need not hold in general game models. But it does hold in full game models, thanks to the existing maximal utility values in rows and columns.

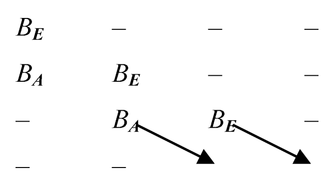
Theorem 6. *Each finite full game model has worlds where Strong Rationality holds.*

Proof. Much as in the proof of Theorem 7, there must be finite *SR-loops* of the form $s_1 \sim_{\boldsymbol{A}} s_2 \sim_{\boldsymbol{E}} \ldots \sim_{\boldsymbol{A}} s_n \sim_{\boldsymbol{E}} s_1$ with $s_1 \models B_{\boldsymbol{E}}, s_2 \models B_{\boldsymbol{A}}, s_3 \models B_{\boldsymbol{E}}, \ldots$ Here is an illustration: each finite full game model has 3-*player SR* loops. In such models, by the earlier observations about maxima on rows and columns, the following assertions are true everywhere:

$$\langle \boldsymbol{B}, \boldsymbol{C}\rangle B_{\boldsymbol{A}}, \langle \boldsymbol{A}, \boldsymbol{C}\rangle B_{\boldsymbol{B}}, \langle \boldsymbol{A}, \boldsymbol{B}\rangle B_{\boldsymbol{C}}.$$

Here the modalities $\langle \boldsymbol{i}, \boldsymbol{j}\rangle$ have an accessibility relation $\sim_{\{i,j\}}$ keeping the co-ordinates for both $\boldsymbol{i}$ and $\boldsymbol{j}$ the same — i.e., the intersection of $\sim_i$ and $\sim_j$. But then, repeating this, by finiteness, we must have loops of the form $B_{\boldsymbol{A}} \sim_{\{\boldsymbol{A},\boldsymbol{C}\}} B_{\boldsymbol{B}} \sim_{\{\boldsymbol{A},\boldsymbol{B}\}} B_{\boldsymbol{C}} \sim_{\{\boldsymbol{B},\boldsymbol{C}\}} B_{\boldsymbol{A}} \sim_{\{\boldsymbol{A},\boldsymbol{C}\}} \ldots$ returning to the initial world with $B_{\boldsymbol{A}}$. Any world in such a loop satisfies $\langle \boldsymbol{A}\rangle B_{\boldsymbol{A}} \wedge \langle \boldsymbol{B}\rangle B_{\boldsymbol{B}} \wedge \langle \boldsymbol{C}\rangle B_{\boldsymbol{C}}$. E.g., if the world itself has $B_{\boldsymbol{A}}$, by reflexivity, it satisfies $\langle \boldsymbol{A}\rangle B_{\boldsymbol{A}}$. Looking back at its mother $B_{\boldsymbol{C}}$ via $\sim_{\{\boldsymbol{B},\boldsymbol{C}\}}$, by symmetry, it has $\langle \boldsymbol{C}\rangle B_{\boldsymbol{C}}$. And looking at its grandmother $B_{\boldsymbol{B}}$ via $\sim_{\{\boldsymbol{B},\boldsymbol{C}\}}$ and $\sim_{\{\boldsymbol{A},\boldsymbol{B}\}}$, by transitivity, it also satisfies $\langle \boldsymbol{B}\rangle B_{\boldsymbol{B}}$. □

Remark 4. Infinite game models.
On infinite game models, *SR*-loops need not occur. Consider a grid of the form *NxN*: Suppose that the best-response pattern runs diagonally as follows:



Then, every sequence $B_{\boldsymbol{E}} \sim_{\boldsymbol{A}} B_A \sim_{\boldsymbol{E}} B_{\boldsymbol{E}} \sim_{\boldsymbol{A}} B_{\boldsymbol{A}} \ldots$ must break off at the top.

But now, we turn to the dynamic epistemic role of these assertions.[9]

5. Iterated Announcement of Rationality and Game Solution

5.1. *Virtual conversation scenarios*

Here is our proposed scenario behind an iterative solution algorithm. We are at some actual world $(\boldsymbol{M}, s)$ in the current game model. Now, players start telling each other things they know about their behaviour at s, narrowing down the available options. As information from another player may change the current game model, it makes sense to iterate the process, and repeat the assertion — if still true. We can take this scenario as real communication, but we see it as *virtual conversation* in the head of individual players. Thus, unlike Muddy Children, which takes place in real time, our game solution scenarios take place in virtual 'reflection time'.

[9]It would be of interest to axiomatize the complete logic of both general and full game models expanded with B_j, B_j^*, WR_j, and SR_j — either in our epistemic base language, or with the technical modal additions of Note 7.

Now, what can players truthfully say? There is of course a trivial solution: just tell the other the action you have chosen. But this is as uninteresting as 'solving' a card game by telling everyone your hand. As with Muddy Children, we look for *generic assertions* that can be formulated in the epistemic language of best response and rationality, without names of concrete actions. And Sec. 4 supplied these.

5.2. *Weak rationality and iterated removal of strictly dominated strategies*

Our first result recasts the usual characterizations of SD^{ω} as follows.

Theorem 7. *The following are equivalent for worlds s in full game models $\boldsymbol{M}(G)$:*

(a) *World s is in the SD^{ω} solution zone of $\boldsymbol{M}(G)$*
(b) *Repeated successive announcement of Weak Rationality for players stabilizes at a submodel $(\boldsymbol{N}, s)$ whose domain is that solution zone.*

Proof. The argument is short, since Sec. 4 has been leading up to this. By its definition, WR_j is true in all worlds which do not lie on a strictly dominated row or column, as the case may be. This argument is easily checked with the alternating update sequence for $WR_E, WR_A, \ldots$ applied to our running Example 1. □

5.3. *The general program*

Theorem 8 is mathematically elementary — and conceptually, it largely restates what we already knew. But the more general point is the style of update analysis as such. We can now match games and epistemic logic in two directions.

> *From games to logic* Given some algorithm defining a solution concept, we can try to find epistemic actions driving its dynamics.

This was the direction of thought illustrated by our analysis of the SD^{ω} algorithm. But there is also a reverse direction:

> *From logic to games* Any type of epistemic assertion defines an iterated solution process which may have independent interest.

In principle, this suggests a general traffic between game theory and logic, going beyond the existing batch of epistemic characterization results which take the game-theoretic solution repertoire as given. Our next illustration shows this potential — though in the end, it turns out to match an existing game-theoretic notion after all.

5.4. *Another scenario: Announcing strong rationality*

Instead of WR, we can also announce Strong Rationality in the preceding scenario. This gives a new game-theoretic solution procedure, whose behaviour can differ.

Example 8. Iterated announcement of *SR*.
Our running example gives exactly the same model sequence as with SD^{ω}:

B_A, B_E	–	–
B_E	B_A	–
–	B_E	B_A

B_A, B_E	–
–	B_A
–	B_E

B_A, B_E	–
B_E	B_A

B_A, B_E
B_E

B_A, B_E

In this particular sequence, a one-world Nash equilibrium model is reached at the end. But *SR* differs from *WR* in this modification of our running example:

		E *a*	*b*	*c*
A	*d*	2, 3	1, 0	1, 1
	e	0, 0	4, 2	1, 1
	f	3, 1	1, 2	2, 1

B_E	–	–
–	B_E, B_A	–
B_A	B_E	B_A

WR does not remove any rows or columns, whereas *SR* removes the top row as well as the right-hand column of this game model.

In general, like *WR*, iterated announcement of *SR* can get stuck in cycles.

Example 9. Ending in *SR*-loops.
In this model, successive announcement of *SR* gets stuck in a 4-cycle:

B_E	–	B_A
–	B_A	B_E
B_A	–	B_E

B_E	B_A
–	B_E
B_A	B_E

B_E	B_A
B_A	B_E

Here is what happens in this update sequence. An individual announcement that j is strongly rational leaves only states s where SR_j is true, making j's rationality *common knowledge*. But this announcement may eliminate worlds from the model, invalidating SR_k at s for *other* players k, as their existential modalities now lack a witness. Thus, announcements of everyone's rationality need not result in common knowledge of joint rationality- and repeated announcement of *SR* makes sense.

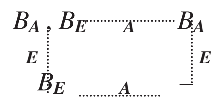
Proposition 4. *Strong Rationality is self-fulfilling on finite full game models.*

Proof. Every finite game model has an *SR loop* (Theorem 7). Worlds in such loops keep satisfying Strong Rationality, and are never eliminated. In finite models, the procedure stops when all worlds have $\sim_{\boldsymbol{E}}$ and $\sim_{\boldsymbol{A}}$ successors on such loops, and Strong Rationality for the whole group has become common knowledge. □

In particular, Nash equilibria present in the game survive into the fixed-point, being *SR*-loops of length 0. But even when these exist, we cannot hope to get just these, as the above description also lets in states with enough links to Nash equilibria.

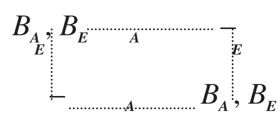
Example 10. Two well-known games.
Prisoner's Dilemma has the following game matrix and epistemic game model:

	E	*a*	*b*
A	*c*	*1, 1*	*3, 0*
	d	*0, 3*	*2,2*



One *SR*-announcement turns this into the one-world Nash equilibrium — as $\langle \boldsymbol{E}\rangle B_{\boldsymbol{E}} \wedge \langle \boldsymbol{A}\rangle B_{\boldsymbol{A}}$ is only true at the top left. Now consider Battle of the Sexes (Example 5):

	E	*a*	*b*
A	*c*	*2,2*	*1,0*
	d	*0,1*	*2,2*



This gets stuck at once in its initial *SR*-loop pattern, as $\langle \boldsymbol{E}\rangle B_{\boldsymbol{E}} \wedge \langle \boldsymbol{A}\rangle B_{\boldsymbol{A}}$ is true everywhere: and repeated announcement of *SR* has no effect at all.

When all is said and done, we have described an existing game-theoretic solution method once more! Iterated announcement of SR amounts to successive removal of actions that are never a best response given the current set of available outcomes. But this is precisely Pearce's game-theoretic algorithm of *Rationalizability*.[10]

5.5. *Comparing iterated WR and SR*

How does the Strong Rationality procedure compare with Weak Rationality, i.e., the standard game-theoretic algorithm SD^{ω}? The assertion *SR* implied *WR* but not vice versa (Sec. 4), and their one-step updates can differ, witness Example 8. But long-term effects of their iterated announcement are less predictable, as *SR* elimination steps move faster than those for *WR*, thereby changing the model.

[10]In line with Apt 2005, there are two options here. Our approach uses a notion of 'best' referring to all actions available in the original model, as codified in our proposition letters. In addition, there is the relative version of 'best response' mentioned in Section 3.3, whose use in iterated announcement of *SR* would rather correspond to Bernheim's version of the Rationalizability algorithm.

E.g., WR is self-fulfilling on any finite general game model, but SR is not, as it fails in some general game models. Nevertheless, we do have a clear connection:

Theorem 8. *For any epistemic model* $\boldsymbol{M}, \#(SR, \boldsymbol{M} \subseteq \#(WR, \boldsymbol{M})$.

This is not obvious, and will only be proved using fixed-point techniques in Sec. 6, which develops the logic of epistemic game models in further detail. Within standard game theory, Pearce has shown that solution procedures based on removing dominated strategies and procedures based on rationalizability yield the same results in suitably rich game models including mixed strategies. Looking at the logical form of corresponding assertions WR and SR, our guess is that this is like validating logical quantifier switches in compact, or otherwise 'completed' models.

5.6. *Other rational things to say*

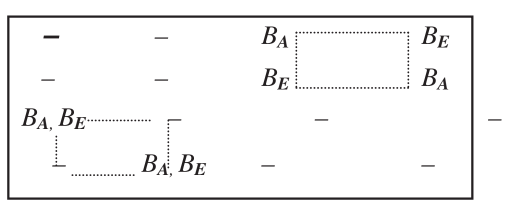
SR is just one new rationality assertion that can drive a game solution algorithm. Many variants are possible in the light of Sec. 4. For instance, let the initial game model have Nash equilibria, and suppose that players have decided on one. The best they can *know* then in the full game model is that they are possibly in such an equilibrium. In this case, we can keep announcing something stronger than SR, viz.

$$\langle \boldsymbol{E} \rangle NE \wedge \langle \boldsymbol{A} \rangle NE \quad \text{Equilibrium Announcement}$$

By the same reasoning as for SR, this is self-fulfilling, and its announcement limit leaves all Nash equilibria plus all worlds which are both $\sim_{\boldsymbol{E}}$ and $\sim_{\boldsymbol{A}}$ related to one.

We conclude with an excursion about possible maximal communication (Sec. 2.5). Once generic rationality statements are exhausted, there may still be ad-hoc things to say, that zoom in further on the actual world, *if* players communicate directly.

Example 11. Getting a bit further.
Consider the following full game model with two different SR-loops, and with the assertion SR true everywhere. The actual world is at the top left, representing some (admittedly suboptimal) pair of decisions for the two players:

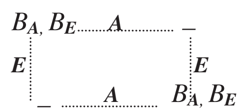


In the actual world, looking down the first column, $\boldsymbol{E}$ knows that $B_{\boldsymbol{A}} \leftrightarrow B_{\boldsymbol{E}}$, so she can announce this. This rules out the third and fourth column. But looking along the first row, $\boldsymbol{A}$ then knows that $\neg B_{\boldsymbol{E}} \wedge \neg B_{\boldsymbol{A}}$, and can announce that. The result is just the top left 4 worlds. These represent common knowledge of $\neg B_{\boldsymbol{E}} \wedge \neg B_{\boldsymbol{A}}$.

There is no epistemic difference between the 4-world model left at the end of Example 11, and a 1-world model with the atomic propositions $B_{\boldsymbol{E}}$, $B_{\boldsymbol{A}}$ both false. For, there is an epistemic *bisimulation* between them (Definition 1), linking all four points to the single one. The same notion tells us when further announcements have no effect. This shows particularly clearly with some basic *SR*-loops, which have already reached the maximal communicative core in the sense of Sec. 2.5.

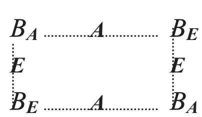
Example 12. Bisimulation contractions of game models.

Consider the two loops in Example 11. The first has two Nash equilibria:



There is an obvious bisimulation between this model and the following one:

Everything is common knowledge here, and no further true epistemic assertion by players can tell one world from another. Next, consider the other *SR*-loop:



This has an obvious epistemic bisimulation with the 2-world model

and the same conclusion applies: we have reached the limit of communication.[11]

If utility values are unique, these two *SR*-loops are the most typical ones to occur.

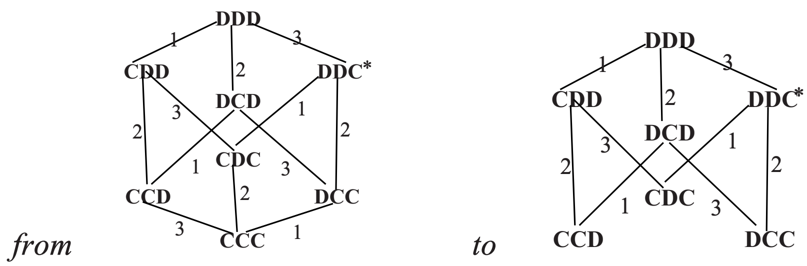
5.7. *Scheduling options*

Finally, the dynamic-epistemic setting has one more degree of freedom in setting up the virtual conversation, viz. its *scheduling*. For instance, the Muddy Children

[11] The final bisimulation contraction takes a game model to a *finite automaton* simulating it. This automata connection to game solution areas may be worth exploring.

of Example 2 had simultaneous announcement of children's knowledge about their status. But its update sequence is quite different if we let the children speak in turn.

Example 2. (*still continued*) Other updates for the Muddy Children.
The first update is as before: $\boldsymbol{CCC}$ disappears:



When the first child says it does not know its status, only world $\boldsymbol{DCC}$ is eliminated. Then in the actual world the second child now knows its status! Saying this takes out all worlds except $\boldsymbol{DDC}$, $\boldsymbol{CDC}$. In that final model, it is common knowledge that $2, 3$ know the truth, while 1 never finds out through pure epistemic assertions.

The same procedural effects might be expected with Strong Rationality instead of joint ignorance. But as we shall see in Corollary 3, SR is less sensitive to order of presentation. Admittedly, first saying $SR_{\boldsymbol{E}}$ and then $SR_{\boldsymbol{A}}$ has different effects from the single $SR_{\boldsymbol{E}} \wedge SR_{\boldsymbol{A}}$. It rather amounts to saying $\langle \boldsymbol{E} \rangle B_{\boldsymbol{E}} \wedge \langle A \rangle (B_{\boldsymbol{A}} \wedge \langle \boldsymbol{E} \rangle B_{\boldsymbol{E}})$. But the latter stronger statement has the same announcement limit as $SR_{\boldsymbol{E}} \wedge SR_{\boldsymbol{A}}$.

A dynamic epistemic approach looks at local effects of sequential assertions. The price for this is order-dependence, and other tricky phenomena from imperative programming. For skeptics, this will be an argument against the approach as such. For fans of dynamics, it just reflects the well-known fact that, in communication and social action generally, matters of procedure crucially affect outcomes.

6. Logical Background: From Epistemic Dynamics to Fixed-Point Logic

6.1. *Issues in dynamic epistemic logic*

Our conversation scenario raises many general issues of dynamic epistemic logic. Some of these are entirely standard ones of *axiomatization*. E.g., with a suitable language including best response, preference comparisons, and rationality assertions, standard epistemic logics of game models encode much of the reasoning in this paper. An example is the existence of SR-loops in full game models in Theorem 7. This can be expressed in epistemic fixed-point logic, as shown in this section. The logic of game models in such a language would be worth determining.

In addition to axiomatization, there are *model-theoretic* issues. A well-known open question in dynamic epistemic logic is the 'Learning Problem' (van Benthem

(2002b)). Some formulas, when announced in a model, always become common knowledge. A typical example are atomic facts, witness the validity of the dynamic-epistemic formula $[p!]C_G p$. Other formulas, when announced, make their own falsity common knowledge. The Moore-style assertion "p, but you don't know it" is a good example: $[(p \wedge \neg K_j p)!]C_G \neg(p \wedge \neg K_j p)$. Yet other formulas make themselves common knowledge only after a finite number of repeated announcements. Or they have no uniformity at all, but become true or false depending on the current model.

Question Exactly which syntactic forms of assertion ϕ have $[\phi!]C_G \phi$ valid?

There are obvious connections with the *self-fulfilling* formulas ψ of Sec. 2, which become common knowledge in their announcement limits $\#(\boldsymbol{M}, \psi)$. E.g., if a formula is uniformly self-fulfilling, being common knowledge in every one of its announcement limits, must it be self-fulfilling after some fixed finite number of steps?

But perhaps the most obvious question is one of computational *complexity*. Let us add announcement limits explicitly to our language:

$$\boldsymbol{M}, s \models \#(\psi) \text{ iff } s \text{ belongs to } \#(\boldsymbol{M}, \psi)$$

Question Is dynamic epistemic logic with # still decidable?

In Sec. 6.4 we show this is true for the special case of $\psi = SR$. Again, this result uses the connection between iterated announcement and epistemic fixed-point logics, providing a more general perspective on our analysis so far.

6.2. *Equilibria and fixed point logic*

To motivate our next step, here is a different take on the original SD^ω algorithm. The original game model itself need not shrink, but we compute a new property of its worlds in approximation stages, starting with the whole domain, and shrinking this until no further change occurs. Such a top-down procedure is like computation of a *greatest fixed point* for some set operator on a domain. Other solution algorithms, such as backward induction, compute *smallest fixed points* with a bottom up procedure. Either way, game solution and equilibrium has to do with fixed points!

Fixed-point operators can be added to various languages, such as standard first-order logic (Moschovakis (1974)). In the present setting, we use an epistemic version of the modal *μ-calculus* (Stirling (1999)). Its semantics works as follows.

Definition 11. Formulas $\phi(p)$ with only positive occurrences of the proposition letter p define the following monotonic set transformation, in any epistemic model $\boldsymbol{M}$:

$$F_\phi(X) = \{s \in \boldsymbol{M} | (\boldsymbol{M}, p := X), s \models \phi\}$$

The formula $\mu p \cdot \phi(p)$ then defines the smallest fixed point of this transformation, starting from the empty set as a first approximation. Likewise, the formula $\nu p \cdot \phi(p)$ defines the greatest fixed point of F_ϕ, starting from the whole domain of $\boldsymbol{M}$ as its first approximation. Both exist for monotone maps, by the Tarski-Knaster theorem.

This is the proper setting for our scenarios in Sec 5. Fixed-point logics work in both finite and infinite models, and hence we obtain full generality. In particular, the *SR*-limit can be defined as a greatest fixed-point in an epistemic μ-calculus:

Theorem 9. *The stable set of worlds for repeated announcement of SR is defined inside the full game model by* $\nu p \cdot (\langle \boldsymbol{E} \rangle (B_{\boldsymbol{E}} \wedge p) \wedge \langle \boldsymbol{A} \rangle (B_{\boldsymbol{A}} \wedge p))$.

Proof. The set of non-eliminated worlds in the *SR* procedure has the right closure properties, and so it is included in the greatest fixed-point. And conversely, no world in the greatest fixed-point is ever eliminated by an announcement of *SR*. □

The equilibrium shows as follows here. The greatest-fixed-point formula $\nu p \cdot (\langle \boldsymbol{E} \rangle (B_{\boldsymbol{E}} \wedge p) \wedge \langle \boldsymbol{A} \rangle (B_{\boldsymbol{A}} \wedge p))$ defines the largest set P from which both agents can see a position which is best for them, and which is again in this very set P.

More precisely, the top-down approximation sequence for any formula $\phi(p)$ looks like this — starting from a formula T true everywhere in the model:

$$T, \quad \phi(T), \quad \phi(\phi(T)), \ldots \text{ taking intersections at limit ordinals}$$

There is a clear correspondence between these stages and elimination rounds in game matrices. Announcing Weak Rationality can be analyzed in a similar fashion.

6.3. *General announcement limits are inflationary fixed points*

There is more to iterated announcement. Recall the announcement limit $\#(\phi, \boldsymbol{M})$ from Sec 5. It arose by continued application of the following function:

Definition 12. Set operator for public announcement.

The function computing the next set for iterated announcement of ϕ is

$$F^*_{M,\phi}(X) = \{s \in X \mid \boldsymbol{M}|X, s \models \phi\}$$

with $\boldsymbol{M}|X$ the restriction of the model $\boldsymbol{M}$ to its subset X.

In general, this function F^* is not monotone with respect to set inclusion, and the epistemic μ-calculus does not apply. The reason was pointed out already in Sec. 3: when $X \subseteq Y$, an epistemic statement ϕ may change its truth value from a model $\boldsymbol{M}|X$ to the larger model $\boldsymbol{M}|Y$. In the same vein, we do not recompute stages in a fixed model, as with formulas $\nu p \cdot \phi(p)$, but in ever smaller ones, changing the range of the modal operators in ϕ all the time. Thus, F^* mixes ordinary fixed-point computation with *model restriction*. But despite the non-monotonicity of its update function, iterated announcement can still be defined in full generality via a broader procedure (Ebbinghaus & Flum (1995)) in so-called *inflationary fixed-point logic*. How this works precisely becomes clear in the proof of the following result.

Theorem 10. *The iterated announcement limit is an inflationary fixed point.*

Proof. Take any ϕ, and relativize it to a fresh proposition letter p, yielding

$$(\phi)^p$$

In the latter formula, p need not occur positively (it becomes negative, e.g., when relativizing positive universal box modalities), and hence a fixed-point operator of the μ-calculus sort is forbidden. An example is

$$(\langle\ \rangle[\ ]q)^p = \langle\ \rangle(p \wedge [\ ](p \to q))$$

Now the Relativization Lemma for logical languages will work with all of $\boldsymbol{M}$. Let P be the denotation of the proposition letter p in $\boldsymbol{M}$. Then for all s in P:

$$\boldsymbol{M}, s \models (\phi)^p \quad \text{iff } \boldsymbol{M}|P, s \models \phi$$

Therefore, the above definition of $F^*_{\boldsymbol{M},\phi}(X)$ as $\{s \in X|\boldsymbol{M}|X, s| = \phi\}$ equals

$$\{s \in \boldsymbol{M}|\boldsymbol{M}[p := X], s \models (\phi)^p\} \cap X$$

But this computes a greatest fixed point of the following generalized sort. Consider any first-order formula $\phi(P)$, without syntactic restrictions on the occurrences of the predicate letter P. Now define an associated map $F^{\#}_{\boldsymbol{M},\phi}(X)$ as follows:

$$F^{\#*}_{M,\phi}(X) = \{s \in \boldsymbol{M}|\boldsymbol{M}[p := X], s \models \phi\} \cap X$$

This map need not be monotone, but it always takes subsets. Thanks to this feature, it can be used to obtain a so-called greatest *inflationary fixed-point* by first applying it to $\boldsymbol{M}$, and then iterating this, taking intersections at limit ordinals. If the function $F^{\#}$ happens to be monotonic, this coincides with the usual fixed point procedure. But general announcement limits for arbitrary ϕ are inflationary fixed points. □

Thus, the general logic of announcement limits can be defined in the known system of inflationary epistemic fixed point logic. In response to a first version of this paper, Dawar, Graedel & Kreutzer (2004) have shown that this is essential: epistemic announcement limits cannot always be defined in a pure μ-calculus. But this insight is also bad news. Modal logic with inflationary fixed points is undecidable, and hence rather complex. Fortunately, special types of epistemic announcement may be better behaved. We show this for our main example of Strong Rationality.

6.4. *Monotone fixed points after all*

Theorem 10 said that iterated announcement of *SR* works via an ordinary greatest fixed-point operator, definable in the epistemic μ-calculus. The reason is that the update function $F_{\boldsymbol{M},SR}(X)$ is indeed monotone for set inclusion. This has to do with the special syntactic form of *SR*, and its model-theoretic preservation behaviour:

Theorem 11. *$F_{M,\phi}(X)$ is monotone for existential modal formulas ϕ.*

Proof. Existential modal formulas are built with only existential modalities, literals, conjunction and disjunction. In particular, no universal knowledge modalities occur. With this syntax, the relativization $(\phi)^p$ has only positive occurrences of p, so F^* is monotone, with an ordinary greatest fixed point computation. □

Existential announcements occur elsewhere, too. Note that this is also the format of the ignorance announcements in the earlier example of the Muddy Children.

Theorem 11 has several applications. The first of these is the earlier Theorem 9 comparing the update sequences for Weak and Strong Rationality:

Corollary 2. *For any epistemic model* $\boldsymbol{M}, \#(SR, \boldsymbol{M}) \subseteq \#(WR, \boldsymbol{M})$.

Proof. By definition, SR implies WR at any world in any general game model. Next, we compare the stages of the fixed-point computation. We always have

$$F^{*\,\alpha}{}_{SR}(\boldsymbol{M}) \subseteq F^{*\,\alpha}{}_{WR}(\boldsymbol{M}) \quad \text{for all ordinal approximations } \alpha$$

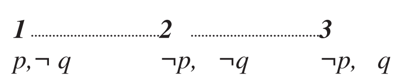
The reason for this is the following inclusion

$$\text{if } X \subseteq Y, \text{ then } F^*_{\boldsymbol{M},SR}(X) \subseteq F^*_{\boldsymbol{M},WR}(Y)$$

This is again a consequence of the special form of our assertions. If $\boldsymbol{M}|X, s \models SR$ and $s \in X$, then $s \in Y$ and so $\boldsymbol{M}|Y, s \models SR$ — by the *existential* form of SR, which is preserved under model extensions. But then also $\boldsymbol{M}|Y, s \models WR$. □

Non-existential forms are more complex. Then, even when ϕ implies ψ in any model $\boldsymbol{M}$, the announcement limit $\#(\phi, \boldsymbol{M})$ need not be included in $\#(\psi, \boldsymbol{M})$!

Example 13. Stronger epistemic formulas may have smaller announcement limits. Consider the pair of formulas $\phi = p \wedge (\langle\rangle\neg p \to \langle\rangle q), \psi = \phi \vee (\neg p \wedge \neg q)$. Now look at this model with accessibility an equivalence relation for a single agent:



The update sequence for ϕ stops in one step with ***1***, while that for ψ runs as follows:

$$\{\mathit{1},\ \mathit{2},\ \mathit{3}\},\ \{\mathit{1},\ \mathit{2}\},\ \{\mathit{2}\}$$

Next, consider the order dependence of Sec. 5.7. Here is why this does not arise in our special case. We do one particular order, but the argument is general.

Corollary 3. *The announcement limit of* $SR_{\boldsymbol{E}}; SR_{\boldsymbol{A}}$ *is the same as that of* SR.

Proof. Sequential announcements $SR_{\boldsymbol{E}}; SR_{\boldsymbol{A}}$ amount to saying $\langle \boldsymbol{E}\rangle B_{\boldsymbol{E}} \wedge \langle A\rangle (B_{\boldsymbol{A}} \wedge \langle \boldsymbol{E}\rangle B_{\boldsymbol{E}})$, as in Sec. 5.7. The latter existential formula implies SR, and so, as in Corollary 2, the announcement limit of $SR_{\boldsymbol{E}}; SR_{\boldsymbol{A}}$ is contained in that of SR.

Conversely, two steps of simultaneous *SR* announcement produce an existential formula implying that for $SR_{\boldsymbol{E}}; SR_{\boldsymbol{A}}$. Hence we also have the opposite inclusion. □

Order independence failed for the case of the Muddy Children. Its driving assertion of ignorance, though existential, involves a negation. Therefore, the single epistemic formula for sequential announcement of ignorance acquires a universal modality. So, its update map is not monotonic, and our argument collapses.

Our final application of Theorem 12 is of a more general logical nature.

Corollary 4. *Dynamic epistemic logic plus* $\#(\psi)$ *for existential* ψ *is decidable.*

Proof. Announcement limits for existential epistemic formulas arise via monotone operators. So they are definable in the epistemic μ-calculus, which is decidable. □

In particular, reasoning about Strong Rationality or Muddy Children stays simple.

6.5. *Greatest fixed points in game generally*

The above suggests a preference for greatest fixed-points in game analysis. Indeed, even bottom-up backward induction can be recast as a top down greatest fixed point procedure. In Zermelo's well-known theorem on determinacy for finite zero-sum two-player games, the node colouring algorithm essentially amounts to evaluating a modal fixed point formula. Van Benthem (2002a) takes a μ-version for the bottom up algorithm, but here is a greatest fixed-point version which works just as well:

$$\nu p \cdot (\text{end} \; \& \; \text{win}_{\boldsymbol{E}}) \vee (\text{turn}_{\boldsymbol{E}} \wedge \langle \varepsilon \rangle p) \vee (\text{turn}_{\boldsymbol{A}} \wedge [\mathfrak{A}] p)$$

This colours every node as a win for player $\boldsymbol{E}$ first — but then, using the universal set as a first approximation, stage by stage, the right colours for $\boldsymbol{A}$ will appear. More generally, strategies seem never-ending resources like our doctors, which can be tapped in case of need, and then return to their original state. This fits well with the recursive character of greatest fixed-points as explained earlier.

7. Richer Models: Worries, External Sources, Beliefs

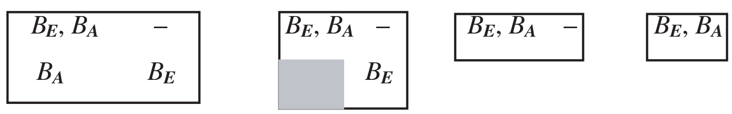
Our proposal makes game solution a process of virtual communication of rationality assertions, resulting in epistemic equilibrium. Of the many possible statements driving this, we have looked only at weak and strong rationality. But the scenario admits of many more variations, some of them already exemplified in Sec. 1.

7.1. *Muddy Children revisited*

The initial information models for Muddy Children are cubes of 3-vectors, which look like full game models. But the self-defeating ignorance assertions suggest an alternative scenario for games, reaching solution zones by repeatedly announcing, not players' rationality, but rather their *worries* that non-optimal outcomes are still a live option. In fact, the story of the Muddy Children as it stands is such a scenario, with actions 'dirty', 'clean'. Children keep saying "my action might turn out well, or badly" — until the first time they know what is in fact the case.

Muddy Children also displays another feature, viz. *enabling actions*. The procedure is jump-started by the Father's initial announcement. Internal communication only reaches the desired goal of common knowledge after some external information has broken the symmetry of the diagram. This also makes sense in games.

Example 13. 'With a little help from my friends'.
Some equilibria may be reached only after external information has removed some strategy profiles, breaking the symmetry of the *SR*-loops of Sec. 8:



The initial model is an *SR*-loop, and nothing gets eliminated by announcing *SR*. But after an initial announcement that, say, the bottom-left world is not a possible outcome, updates take the resulting 3-world model to its single Nash equilibrium.

Every equilibrium world or solution zone can be obtained in this way, if definable in our language. The art is to find *plausible* external announcements which can set the virtual conversation going, or intervene at intermediate stages.

7.2. *Changing beliefs and plausibility*

The epistemic game models $(\boldsymbol{M}, s)$ of Sec. 3 with just relations $\sim_j$ may seem naive. Players are supposed to know their own action already — while this is precisely what they are trying to choose through deliberation! A more delicate analysis of players' attitudes in solution procedures would need at least *beliefs*. Stalnaker (1999) has sophisticated models of this sort. Fortunately, just to illustrate our dynamic stance, beliefs can be analyzed in a simple manner with world-eliminating update procedures. Many standard logics of belief enrich epistemic models with orderings $\leq_j$ of *relative plausibility* among those worlds which agent j cannot epistemically distinguish. Belief by an agent is then truth in all her most plausible alternatives:

$$\boldsymbol{M}, s \models B_j\phi \quad \text{iff for all } \leq_j\text{-}\textit{best} \text{ worlds in } \{t \mid t \sim_j s\} : \boldsymbol{M}, t \models \phi$$

This is less demanding than the earlier uncertainty semantics for knowledge: e.g., beliefs can be false when the actual world is not $\leq_j$-best. The plausibility order also supports other logical operators, such as an agent-dependent *conditional*:

$$\boldsymbol{M}, s \models \phi \Rightarrow_j \psi \qquad \text{iff for all } \leq_j \text{-best worlds in } \{t \mid \boldsymbol{M}, t \models \phi\} : \boldsymbol{M}, t \models \psi$$

There are again dynamic-doxastic reduction axioms like those for knowledge. Van Benthem (2002d) notes that after a public announcement the resulting beliefs satisfy

$$[A!]B_j\phi \leftrightarrow (A \Rightarrow_j [A!]\phi)$$

$$[A!]\phi \Rightarrow_j \psi \leftrightarrow ((A \wedge [A!]\phi) \Rightarrow_j [A!]\psi)$$

Thus a conditional encodes right now what agents would believe when updated. Beyond world elimination by public announcements, recent dynamic logics also describe how agents' *plausibility relations change* as new information comes in, generalizing existing belief revision theories (Aucher (2003), van Benthem (2005)).

Example 14. Updating plausibilities.
Consider our running Example 1 again, now with all worlds equally plausible for all agents in the initial model. Another type of conversation scenario might use 'soft updates' that modify these expectations. The trigger might be rationality assertions

> If j believes that a is a possible action, and b is always worse than a (in terms of j's preferences among outcomes) with respect to actions of the other player that j considers possible, then j does not play b.

As a soft update, such an implication does not eliminate worlds with b in them, but it makes them *less plausible* than all others. Thus, worlds in the discarded columns and rows of the SD^ω algorithm lose plausibility, making them irrelevant for the new beliefs after the update. The result of the sequence of plausibility relations would make players believe they are in the solution set of the algorithm.

This is just one of many ways in which our dynamic epistemic analysis of game solution procedures can be refined to include beliefs, as well as other update triggers.

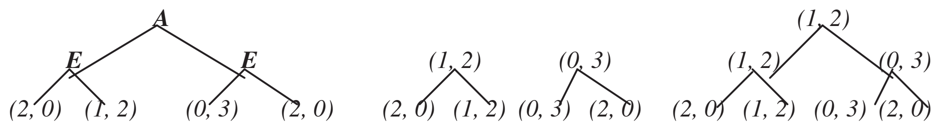
8. A Test Case: Epistemic Procedures in Extensive Games

Our announcement scenarios worked on strategic games. But *extensive games* are no obstacle. A solution algorithm like Backward Induction suggests similar epistemic procedures. As in Sec. 3, we first need to decide on models to work with. The literature often has worlds including complete strategy profiles ω as before, with some added game node s. But sometimes, this approach seems overly structured, and we can stay closer to the game tree of an extensive game, interpreting some standard *branching-temporal language* (cf. van Benthem (2002d), van Benthem, van Otterloo & Roy (2005)). At nodes of the tree, players still see a set of possible histories continuing the one so far. Further information may then lead them to rule out branches from this set.

8.1. *Backward Induction analyzed*

First, consider a very simple standard case of the procedure.

Example 15. Backward Induction.
Here are the successive steps computing node values in a simple case:



This is a bottom-up computation for node values. But we can also recast it as an elimination procedure for branches, driven by iterated announcement of an analogue of the earlier rationality principles of Sec. 4. Here is one version:

Definition 13. Momentaneous Rationality.
Momentaneous rationality MR says that at every stage of a branch in the current model, the player whose turn it is, has not selected a move whose available continuations all end worse for her than all those after some other possible move.

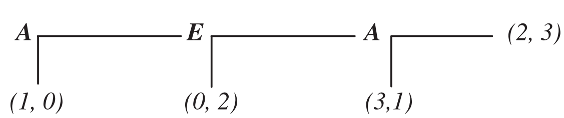
Announcing *MR* removes at least those histories from the game tree which would be deleted by one backward induction step. Moreover, repeated announcement makes sense, as a smaller bundle of possible future histories may trigger new eliminations. Sometimes, the *MR* process may go faster than backward induction. In Example 16, both rightmost branches would be eliminated straightaway by announcing *MR* if the value of the right end node $(2, 0)$ had been $(1/2, 0)$. But the end result is the same:

Proposition 5. *On finite extensive game trees, iterated announcement of MR arrives exactly at the Backward Induction solution.*

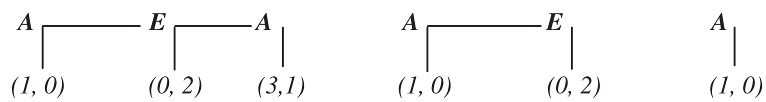
Again, this announcement scenario also suggests alternative solution procedures. For instance, a more co-operative scenario might involve an assertion of

> *Cooperative Rationality CR*
> Players never select a move m when there is another move allowing at least one outcome better for both players than any history following m.

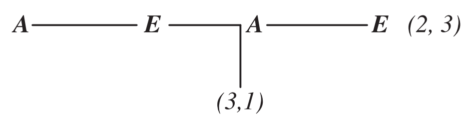
Example 16. *MR* and *CR* conversation in a Centipede game.
Here is a famous example, of which we just show a simple case:



Backward induction computes value $(1, 0)$ for the initial node: ***A*** plays down. Iterated announcement of the assertion *MR* does this in three stages:



This is the controversial outcome where players would be better off going to the end, where ***A*** gets more than 1, and ***E*** more than 0. The above co-operative announcement proposal, however, would indeed make a different prediction. Iterated announcement of *CR* first rules out the first down move for ***A***, and then the following down move for ***E***. After that it leaves both options for ***A*** at the end.



Further announcements might enforce a unique solution, unlike *CR* by itself. One of these might be "I will repay favours", i.e., the risks of losing some guaranteed amount that you have run on behalf of a better outcome for both of us. Thus, we can choose models, languages, and procedures for extensive games driving the same scenarios as in our analysis of strategic games — with even more options.[12]

9. Conclusion

Dynamic intuitions concerning activities of deliberation and communication lie behind much of epistemic logic and related themes in game theory — though they are often implicit. In physics, an equilibrium is only intelligible if we also give an explicit dynamic account of the forces leading to it. Likewise, epistemic equilibrium is best understood with an explicit logical account of the actions leading to it. For this purpose, we used update scenarios for scenarios of virtual communication, in a dynamic epistemic logic for changing game models. This new stance also fits better with our intuitive term *rationality*. One sometimes talks about rational outcomes, which satisfy some sort of harmony between utilities and expectations. But the more fundamental notion may be that of rational agents performing *rational actions*. Taken in the latter sense, our rationality is located precisely in the procedure being followed.

Summarizing our main findings, solving a game involves dynamic epistemic procedures which are of interest per se, and game-theoretic equilibria are their

[12]It is tempting now to consider extensive games as they evolve over time. Players then experience two different process, viz. *update with observed moves* plus *revision of expectations* about the future course of the game. To do justice to this, we need a more complex dynamic-epistemic-temporal logic. Also, more complex global hypotheses about behaviour than *MR* or *CR* (say, 'you are a finite automaton') take us to full-fledged strategy-profile worlds after all. van Benthem (2002d) has further discussion and a richer temporal framework for dealing with such scenarios.

greatest fixed points. This analogy suggests a general study of game solution concepts in dynamic epistemic logic, instead of just separate epistemic characterization theorems. Sections 5 and 6 identified a number of model-theoretic results on dynamic epistemic logics which show there is content to this. In particular, game-theoretic equilibrium got linked to computational fixed-point logics, which have a sophisticated theory of their own that may be useful here. But mainly, we hope our scenarios are just fun to explore, extend, and generally: play with!

Finally, our analysis has obvious limitations. The models are crude, and cannot make sophisticated epistemic distinctions. Moreover, we have ignored the role of probability and mixed strategies throughout. We certainly do not claim that explicit epistemic dynamics is a miracle cure for the known cracks in the foundations of game theory. But it does add a new way of looking at things, as well as one more sample of promising contacts between games, logic, and computation.

Acknowledgement

I thank Krzysztof Apt, Boudewijn de Bruin, Martin Hyland, and two anonymous referees of this journal for their comments.

References

Aucher, G. [2003] *A Joint System of Update Logic and Belief Revision* (Master of Logic Thesis, ILLC, Amsterdam).

Baltag, A., Moss, L. and Solecki, S. [1998] The logic of public announcements, common knowledge and private suspicions, *Proceedings TARK* (Morgan Kaufmann Publishers, Los Altos) (Updated versions through 2004), 43–56.

van Benthem, J. [1996] *Exploring Logical Dynamics* (CSLI Publications, Stanford).

van Benthem, J. [2002a] Extensive games as process models, *Journal of Logic, Language and Information* **11**, 289–313.

van Benthem, J. [2002b] One is a lonely number: On the logic of communication, Colloquium Logicum, Muenster. Tech Report PP-2002-27, ILLC Amsterdam, in Z. Chatzidakis, P. Keopke & W. Pohlers (eds.), *Logic Colloquium '02*, ASL & A.K. Peters, Wellesley MA, pp. 96–129.

van Benthem, J. [2002c] Rational dynamics and epistemic logic in games, Tech Report PP-2002-26, ILLC Amsterdam.

van Benthem, J. [2002d] Update and revision in games, Working paper, ILLC Amsterdam & Stanford University (Various versions through 2004).

van Benthem, J. [2005] Dynamic logic of belief revision, ILLC Tech Report, Amsterdam, to appear in *Journal of Applied Non-Classical Logics*.

van Benthem, J., van Eijck, J. and Kooi, B. [2004] Axioms for communication and action (ILLC, Amsterdam, CWI Amsterdam & Philosophy, Groningen) in *Information and Computation* **204**(11) (2006) 1620–1662.

van Benthem, J. van Otterloo, S. and Roy, O. [2005] Preference logic, conditionals, and solution concepts in games, manuscript, ILLC Amsterdam, in Langerlund, H. Lindström, S.& Sliwinski, R. (eds.) *Modality Matters*, University of Uppsala, 61–76.

Blackburn, P., de Rijke, M. and Venema, Y. [2000] *Modal Logic* (Cambridge University Press, Cambridge).

de Bruin, B. [2003] *Explaining games—On the Logic of Game-Theoretic Explanations*, Dissertation DS 2004-03, ILLC, University of Amsterdam.

Binmore, K. [1992] *Fun and Games*, Heath & Company, Lexington (Mass.).

Dawar, A., Grädel, E. and Kreutzer, S. [2004] Inflationary fixed points in modal logic, *ACM Transations on Computational Logic* **5**, 282–315.

van Ditmarsch, H. [2002] Keeping secrets with public communication Department of Computer Science, University of Otago, Dunedin.

van Ditmarsch, H., van der Hoek, W. and Kooi, B. [2004] Dynamic epistemic logic, to appear with Kluwer-Springer Publishers, Dordrecht (2007).

Ebbinghaus, H.-D. and Flum, J. [1995] *Finite Model Theory* (Springer, Berlin).

Fagin, R., Halpern, J., Moses, Y. and Vardi, M. [1995] *Reasoning about Knowledge*, MIT Press, Cambridge (Mass.).

Meijer, J.-J. and van der Hoek, W. [1995] *Epistemic Logic for AI and Computer Science* Cambridge University Press, Cambridge.

Moschovakis, Y. [1974] *Elementary Induction on Abstract Structures*, North-Holland, Amsterdam.

Osborne, M. and Rubinstein, A. [1994] *A Course in Game Theory*, MIT Press, Cambridge (Mass.).

Stalnaker, R. [1999] Extensive and strategic form: Games and models for games, *Research in Economics* **53**, 293–291.

Stirling, C. [1999] Bisimulation, modal logic, and model checking games, *Logic Journal of the IGPL* **7**(1), 103–124.