RATIONAL DYNAMICS AND EPISTEMIC LOGIC IN GAMES

Johan van Benthem, Amsterdam & Stanford

Revised Version, February 2006

Abstract

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.

1 Reaching equilibrium as an epistemic process

1.1 Inductive solution algorithms for games

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

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 instruction. 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, leading to successive eliminations leaving 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 can perform 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 a given solution concept are just those that occur in epistemic models for some suitable rationality assertion involving mutual knowledge and beliefs of players. De Bruin 2004 has a survey of the mathematics of twenty years of such 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: solving Muddy Children by 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:* such as 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' in the 1980s, which also shows in belief revision theories in AI, linear 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 all see the others, but not themselves, so they do not know their own status. Now Father comes and 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?

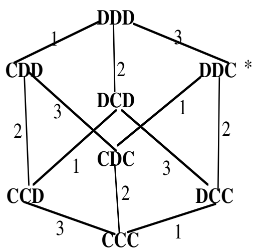
> Nobody knows in the first round. But in the next round, each muddy child can reason like this: “If I were clean, the one dirty child I see would have seen only clean children around her, and so she would have known that she was dirty at once. But she did not. So I must be dirty, too!” ♣

In the initial epistemic model for this situation, eight possible worlds assign ***D*** or ***C*** to each child. A child knows about the others’ faces, but not about his 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.

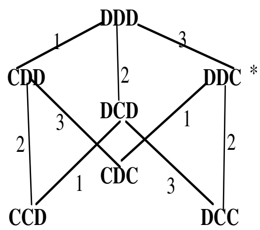
Example 2, continued Updates for the muddy children.

Updates start with the Father's public announcement that at least one child is dirty. This is about the simplest communicative action, and it merely eliminates those worlds from the initial model where the stated proposition is false. I.e., ***CCC*** disappears:

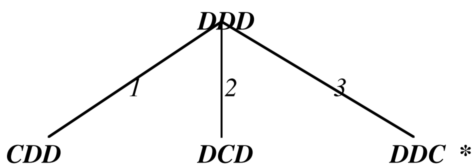
from



to



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



The final update is to ***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 ***D*** and ***C*** for the whole group.

Note that 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 greater 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:

1	*2*	*3*
4	*5*	*6*
7	*8*	*9*

1	*2*
4	*5*
7	*8*

1	*2*
4	*5*

1
4

1

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 Section 3, we shall see *which assertion* drives these elimination steps. Analyzing the algorithm in this detailed manner involves both epistemic logic and dynamic action logic. In particular, in modern dynamic-epistemic logics (cf. Section 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 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. First, in Section 2, we explain the machinery of dynamic-epistemic logic, including the intriguing behaviour of repeated assertions. Section 3 has the epistemic game models with preference structure that we will work with, and we explore their logic. In Section 4, we define two major options for 'rationality' of players, and find a complete description of the finite models of interest. Then in Section 5, we analyze SD^{ω} as repeated assertion of 'weak rationality', and also show how linking solution algorithms with announcement procedures suggests a variant algorithm driven by 'strong rationality' which matches Rationalizability. Section 6 analyzes the resulting framework for game analysis in general terms, 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, Section 7 points out further relevant features of the Muddy Children puzzle scenario, considering also epistemic procedures that revise players' beliefs. Section 8 discusses generalizations of our dynamic-epistemic style of 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'):

$$p \mid \neg\phi \mid \phi \lor \psi \mid K_i\phi \mid C_G\phi$$

We also write $<j>\phi$ for the dual modality of truth in at least one accessible world. This paper uses standard multi-*S5* models ***M*** whose accessibilities are equivalence relations for agents. We write *(**M**, s)* for models with a current world *s*, suppressing brackets when convenient. Also, all models are *finite*, unless otherwise specified.

The key semantic clauses are as follows:

$\boldsymbol{M}, s \models K_i\phi$ iff for all t with $s \sim_i t$, $\boldsymbol{M}: t \models \phi$

$\boldsymbol{M}, s \models C_G\phi$ iff for all t that are reachable from s by some finite sequence of $\sim_i$ steps (any $i \in G$): $\boldsymbol{M}, t \models \phi$

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 informational situation from the viewpoint of our standard epistemic language.

Definition 1 Epistemic bisimulation.

A *bisimulation* between epistemic models ***M***, ***N*** is a binary relation $\equiv$ between states *m, n* in ***M***, ***N*** such that, whenever $m \equiv n$, then (a) *m, n* satisfy the same proposition letters, (b1) if $m\ R\ m'$, then there exists a world n' with $n\ R\ n'$ and $m' \equiv n'$, (b2) the same ‘zigzag clause’ holds in the opposite direction. ♣

Every model *(**M**, s)* has a smallest bisimilar model *(**N**, s)*, its 'bisimulation contraction'. The latter is the simplest representation of the epistemic information in *(**M**, s)*. Next, the following connection with our language is easily proved by formula induction.

Proposition 1 Invariance for bisimulation.

For every bisimulation E between two models ***M***, ***N*** with $s\ E\ t$: 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.

Sometimes, we also need *distributed knowledge* in a group G:

$\boldsymbol{M}, s \models D_G\phi$ iff for all t with $s \bigcap_{i\in G} \sim_i t$: $\boldsymbol{M}, t \models \phi$ [1] ♣

2.2 Public announcement and model change by world elimination

Now we extend this language to also describes 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:

$$\boxed{\boldsymbol{p} \overset{1}{\cdots\cdots\cdots\cdots} \neg p}$$

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

$$\boxed{\boldsymbol{p}}$$

where p has become common knowledge in the group *{1, 2}*. ♣

---

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

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 $\{t \in \boldsymbol{M} \mid \boldsymbol{M}, t \models P\}$:

from			*to*		
$\boldsymbol{M}, s$	s P	$\neg P$	$\boldsymbol{M}|P$, s	s	

♣

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 eventually as worlds drop out. Thus, in the end, common knowledge was achieved.

2.3 Public announcement logic

Update can be studied 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

$[P!]\phi$ after truthful public announcement of P, formula ϕ 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:

$[P!]q$	$\leftrightarrow$	$P \rightarrow q$	*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 must stop in finite models. 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, we have a model where ϕ has become common knowledge. We call such statements *self-fulfilling* in the given model, all others are *self-refuting*. ♣

The rationality assertions for games in Section 4 are self-fulfilling. The joint ignorance statement of the muddy children was self-refuting: inducing common knowledge of its negation. Thus, both types of announcement limit can be of positive interest. [2]

2.5 Maximal group communication

If the muddy children tell each other what they see, common knowledge of the actual world is reached at once. We now describe what a group can achieve by maximal public announcement. Epistemic 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 $COM(\boldsymbol{M}, s) = \{t \mid s \cap_{i \in G} \sim_i t\}$.

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. Now, 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 $\{t \mid s \sim_i t\}$ whose defining proposition is known to agent i, the agent could state this, as he knows it. 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. ♣

[2] There is also 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 announcement limit consists of all worlds that could have been actual in this way. With this said, we mostly suppress mention of s.

[3] *COM(**M**, s)* is also 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 exist describing partial observation, hiding, misleading, and cheating. but they are not needed for 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 presupposes a choice of static epistemic game models serving as the 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.

3.1 Epistemic game models

Consider a strategic game $\boldsymbol{G} = (I, \{A_j \mid j \in I\}, \{P_j \mid j \in I\})$ 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*, each of which determines a unique outcome – and each player has his own

[3] In fact, agents can reach *COM(**M**, s)* by speaking just once. Over-all, we interleave bisimulation contractions with update, making sure all subsets are epistemically definable. First, agent *1* communicates all he knows by stating the disjunction $\bigvee\delta_\tau$ for all worlds t he considers indistinguishable from the actual one. This initial move cuts the model down to the set $\{t \mid s \sim_1 t\}$. 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 $\{t \mid s \sim_1 t\}$ down to those worlds that are also $\sim_2$-accessible from the actual one. Repeating this leads to the submodel *COM(**M**, s),*

[4] More delicate planning includes announcing facts publicly between some agents while leaving other group members in the dark about the actual world (cf. the 'Moscow Puzzle' in van Ditmarsch 2002).

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 structure $\boldsymbol{M(G)}$ whose worlds are all strategy profiles, and whose epistemic accessibility $\sim_j$ for player j is defined as 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 the 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 here. 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.

Example 4 Matrix game models.

The model for the matrix game in Example 1 looks as follows:

$$\begin{tikzcd}
(d, a) \arrow[r, dash, dotted, "A"'] \arrow[d, dash, dotted, "E"] & (d, b) \arrow[r, dash, dotted, "A"'] \arrow[d, dash, dotted, "E"] & (d, c) \arrow[d, dash, dotted, "E"] \\
(e, a) \arrow[r, dash, dotted, "A"'] \arrow[d, dash, dotted, "E"] & (e, b) \arrow[r, dash, dotted, "A"'] \arrow[d, dash, dotted, "E"] & (e, c) \arrow[d, dash, dotted, "E"] \\
(f, a) \arrow[r, dash, dotted, "A"'] & (f, b) \arrow[r, dash, dotted, "A"'] & (f, c)
\end{tikzcd}$$

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

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

Definition 6 A *general game model* $\boldsymbol{M}$ 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

Full or general game models support any type of epistemic statement in the languages of Section 2.1. Some examples will follow in Section 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* $K_AK_E\phi \rightarrow K_EK_A\phi$. 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 modal *Grzegorczyk Axiom* for the modality K_AK_E (cf. Blackburn, de Rijke & Venema 2000). Such principles are crucial to modal reasoning about players' epistemic situations. [5]

3.3 Best response and Nash equilibria

To talk about solutions and equilibria we need some 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

[5] Actually, as to computational complexity, the logic of arbitrary full game models (finite or infinte) 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.

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:

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

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

Stated in this way, best response is an *absolute* property whose conjunction runs over 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:

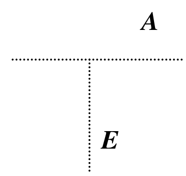
	E	*a*	*b*
A	*c*	*2, 1*	*0, 0*
	d	*0, 0*	*1, 2*

$$\begin{tikzcd} (c, a) \arrow[r, dash, dotted, "\boldsymbol{A}"] \arrow[d, dash, dotted, "\boldsymbol{E}"'] & (c, b) \arrow[d, dash, dotted, "\boldsymbol{E}"] \\ (d, a) \arrow[r, dash, dotted, "\boldsymbol{A}"] & (d, b) \end{tikzcd}$$

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 distinction. B_j says 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: $\langle E \rangle B_A \wedge \langle A \rangle B_E$. It expresses the final observation of Example 5. We will see further valid principles later on. ♣

But there are also alternative logical 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 ***M*** looks only at the strategy profiles available *inside* ***M***. After all, in that model players know that these are the only action patterns that will occur.

Definition 8 Relative best response.

The *relative best response* proposition B^*_j in a general game model ***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 ***M***. ♣

With B^*_j, best profiles for *j* may change as the model changes. For instance, 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]

---

[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.

Remark 2 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 about *j* 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 Section 4. Finally, the connection: absolute-best implies relative-best, but not vice versa.

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 is playing 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?

[7] The original version of this paper (van Benthem 2002C) has a richer language for game models with preference modalities, nominals for specific worlds, a universal modality over worlds, and distributed group knowledge. These modal gadgets yield formulas like $K_A{\downarrow}\ \langle E\rangle\ act_E \geq_E {\downarrow}act_E$ which formalize various 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.

This issue is also important in our epistemic conversation scenarios for game solution (Sections 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, even if it is – 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

$$WR_j \qquad \&_{a\neq\omega(j)} <j> \text{'}j\text{'s current action is at least as good for } j \text{ as } a\text{'}$$

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

The Weak Rationality assertion WR_j has been defined to fail exactly at those rows or columns in a two-player general game model that are strictly dominated for j. For instance, 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 WR_j true for all players j.

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

$$s_1 \sim_A s_2 \sim_E \ldots \sim_A s_n \sim_E s_1 \qquad \text{with } s_1 \models B^*_E,\ s_2 \models B^*_A,\ s_3 \models B^*_E,\ \ldots$$

[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.

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. Section 3):

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

E.g., the first says that, given a world with some action for $\boldsymbol{E}$, there must be some 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 $B^*_E \sim_E B^*_A \sim_A B^*_E \sim_E B^*_A$ … which must loop since the model is finite. Thus, some world in the sequence with, say, B^*_A must be $\sim_A$ –connected to some earlier world *w*. Now, either *w* has B^*_E, or *w* has a successor with B^*_E via $\sim_A$ in the sequence. The former case reduces to the latter by the transitivity of $\sim_A$. 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: $\langle E \rangle B^*_E \wedge \langle A \rangle B^*_A$. ♣

Proposition 2

Weak Rationality is epistemically introspective.

Proof By Definition 9 and the 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 epistemic 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 expanded 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 $\langle j\rangle\phi \rightarrow K_j\langle j\rangle\phi$, 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 which is 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 5, 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. It says that

> *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 (cf. de Bruin 2004, Apt 2005), where one discards actions for which a better response exists under all circumstances. We will return to this later.

Strong Rationality need not be satisfied 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 are *SR–loops* of the form

$$s_1 \sim_A s_2 \sim_E \ldots \sim_A s_n \sim_E s_1 \text{ with } s_1 \models B_E,\ s_2 \models B_A,\ s_3 \models B_E,\ \ldots$$

Here we give an extended 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 is true everywhere:

$$\langle \boldsymbol{B}, \boldsymbol{C}\rangle B_A,\ \langle \boldsymbol{A}, \boldsymbol{C}\rangle B_B,\ \langle \boldsymbol{A}, \boldsymbol{B}\rangle B_C$$

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 $B_A \sim_{\{A, C\}} B_B \sim_{\{A, B\}} B_C \sim_{\{B, C\}} B_A \sim_{\{A, C\}} \ldots$ returning to the initial world with B_A. Any world in such a loop satisfies $\langle \boldsymbol{A}\rangle B_A \wedge \langle \boldsymbol{B}\rangle B_B \wedge \langle \boldsymbol{C}\rangle B_C$. E.g., if the world itself has B_A, by reflexivity, it satisfies $\langle \boldsymbol{A}\rangle B_A$. Looking back at its mother B_C via $\sim_{\{B, C\}}$, by symmetry, it has $\langle \boldsymbol{C}\rangle B_C$. And looking at its grandmother B_B via $\sim_{\{B, C\}}$ and $\sim_{\{A, B\}}$, by transitivity, it also satisfies $\langle \boldsymbol{B}\rangle B_B$. ♣

Remark 3 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:

B_E	–	–	–
B_A	B_E	–	–
–	B_A	B_E	–
–	–		

Then, every sequence $B_E \sim_A B_A \sim_E B_E \sim_A B_A$ … 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 *(**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. Moreover, 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 our preference is 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'.

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. On the analogy of Muddy Children, we look for

[9] It would be of interest to axiomatize the logic of 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 additions of Note 7.

generic assertions that can be formulated in the epistemic language of best response and rationality, without names of concrete actions. And Section 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 Section 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, ... 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 is going on 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. For the same reason, announcements of everyone's rationality need not result in common knowledge of joint rationality. Thus, repeated announcement of SR makes sense.

Proposition 4

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

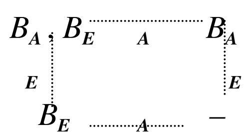
Proof Every finite game model has an *SR loop* (Theorem 7). Worlds in such loops keep satisfying Strong Rationality, and are never eliminated. In finite models, the procedure stops when all worlds have $\sim_E$ and $\sim_A$ successors on such loops, and Strong Rationality for the whole group has become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

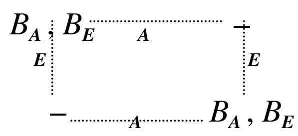
Prisoner's Dilemma has the following game matrix and epistemic game model:

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



One SR-announcement turns this into the single world Nash equilibrium – as $\langle E\rangle B_E \wedge \langle A\rangle B_A$ is only true at the top left. Now consider Battle of the Sexes (cf. 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 $<\boldsymbol{E}>B_E \wedge <\boldsymbol{A}>B_A$ is true everywhere: and repeated announcement of *SR* has no effect at all. ♣

But 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 procedure is precisely Pearce's game-theoretic algorithm of *Rationalizability*. [10]

5.5 Comparing iterated WR and SR

How does the Strong Rationality elimination procedure compare to that with Weak Rationality, i.e., the standard game-theoretic algorithm SD^{ω}? The assertion *SR* implied *WR* but not vice versa (Section 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. For instance, *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 shown using fixed-point techniques in Section 6, which develops the logic of epistemic game models in further detail. Again 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 the corresponding assertions *WR* and *SR*, our guess is that this is like validating logical quantifier switches in compact, or otherwise 'completed' models.

[10] In line with Apt 2005, we note that there are really 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.

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 Section 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.

$$\langle \boldsymbol{E} \rangle\, NE \wedge \langle \boldsymbol{A} \rangle\, NE \qquad\qquad \textit{Equilibrium Announcement}$$

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 $\sim_E$ and $\sim_A$ related to one.

We conclude with an excursion about possible maximal communication (Section 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:

–	–	B_A	B_E
–	–	B_E	B_A
$B_{A,} B_E$	–	–	–
–	$B_{A,} B_E$	–	–

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

There is no epistemic difference between the *4*-world model left at the end of Example 11, and just a *1*-world model with the atomic propositions B_E, B_A both false. For, there is an epistemic *bisimulation* between them, in the sense of 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 Section 2.5.

Example 12 Bisimulation contractions of game models.

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

$$\begin{tikzcd} B_{A,} B_E \arrow[r, dash, dotted, "A"] \arrow[d, dash, dotted, "E"'] & - \arrow[d, dash, dotted, "E"] \\ - \arrow[r, dash, dotted, "A"] & B_{A,} B_E \end{tikzcd}$$

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

$$\begin{tikzcd} B_{A,} B_E \arrow[r, dash, dotted, "{A, E}"] & - \end{tikzcd}$$

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

$$\begin{tikzcd} B_A \arrow[r, dash, dotted, "A"] \arrow[d, dash, dotted, "E"'] & B_E \arrow[d, dash, dotted, "E"] \\ B_E \arrow[r, dash, dotted, "A"] & B_A \end{tikzcd}$$

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

$$\begin{tikzcd} B_A \arrow[r, dash, dotted, "{A, E}"] & B_E \end{tikzcd}$$

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

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

[11] The final bisimulation contraction takes a game model to a simple *finite automaton* simulating it, an 'irreducible outcome type'. This automata connection to game solution areas may be worth exploring.

Even so, at this early stage, we have not been able to find really convincing further epistemic assertions that would drive truly new game-theoretic solution procedures.

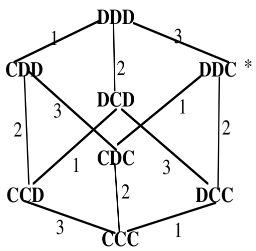
5.7 Scheduling options

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 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.

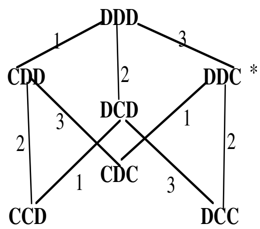
Example 2, still continued Other updates for the Muddy Children.

The first update is as before: ***CCC*** disappears:

from



to



When the first child says it does not know its status, only world ***DCC*** is eliminated. Then in the actual world the second child now knows its status! Saying this eliminates all worlds except ***DDC***, ***CDC***. In the resulting 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 the children's joint ignorance. But as we shall prove in Corollary 3, *SR* is less sensitive to order of presentation. Admittedly, first saying SR_E and then SR_A has different effects from the single $SR_E \wedge SR_A$. It rather amounts to saying $\langle \boldsymbol{E} \rangle B_E \wedge \langle A \rangle (B_A \wedge \langle \boldsymbol{E} \rangle B_E)$. But the latter stronger statement has the same announcement limit as $SR_E \wedge SR_A$.

Even so, a dynamic epistemic approach looks at local effects of sequential assertions, and the price for this is order-dependence, and other tricky phenomena known 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, and hence the complete 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 Section 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:

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

Question Is dynamic epistemic logic with # still decidable?

In Section 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 logical 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 $\boldsymbol{M}$:

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

The formula $\mu p\bullet\ \phi\ (p)$ then defines the smallest fixed point of this transformation, starting from the empty set as a first approximation. Likewise, the formula $\nu p\bullet\ \phi\ (p)$ 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 Section 5. 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 $\nu p \bullet (<\boldsymbol{E}>(B_E \wedge p) \wedge <\boldsymbol{A}>(B_A \wedge p))$.

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

The equilibrium character shows as follows in this format. The greatest-fixed-point formula $\nu p \bullet (<\boldsymbol{E}>(B_E \wedge p) \wedge <\boldsymbol{A}>(B_A \wedge p))$ 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:

$T, \qquad \phi(T), \qquad \phi(\phi(T)), \qquad \ldots$ 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

But there is more to iterated announcement. Recall the definition of the announcement limit $\#(\phi, \boldsymbol{M})$ in Section 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 analysis in the epistemic μ–calculus does not apply. The reason was pointed out already in

Section 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 $\nu p\bullet\ \phi(p)$, 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 sort of 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

$$(\langle\rangle[\,]q)^p \;=\; \langle\rangle(p \wedge [\,](p \rightarrow q))$$

Now the Relativization Lemma for logical languages can be applied to 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:

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

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

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

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:

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

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 $F_{M,\, SR}(X)$ *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 special syntax, the above 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 their definitions, *SR* implies *WR* at any world in any general game model. Next, we compare the stages of the fixed-point computation. We always have that

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

The reason for this is the following inclusion

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

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 also $\boldsymbol{M}|Y,\ s \models SR$ – by the *existential* definition of *SR*, which makes it preserved under model extensions. But then also $\boldsymbol{M}|Y,\ s \models WR$. ♣

The situation with non-existential forms is 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 $\phi = p \wedge (\langle\rangle\neg p \rightarrow \langle\rangle q)$, $\psi = \phi \vee (\neg p \wedge \neg q)$. Now look at this model with accessibility just an equivalence relation for a single agent:

1 ·········	***2*** ·········	***3***
$p, \neg q$	$\neg p, \neg q$	$\neg p, q$

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

$\{\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}\}, \{\boldsymbol{1}, \boldsymbol{2}\}, \{\boldsymbol{2}\}$ ♣

Next, we consider the potential order dependence of Section 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 $SR_E\ ;\ SR_A$ is the same as that of *SR*.

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

Conversely, two steps of simultaneous *SR* announcement also produce an existential formula implying that for SR_E ; SR_A. Hence we also have the opposite inclusion. ♣

This order independence failed for the case of the Muddy Children. The reason is that 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 with $\#(\psi)$ added 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. E.g., 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:

$$\nu p \bullet (end \,\&\, win_E) \vee (turn_E \wedge \langle E \rangle p) \vee (turn_A \wedge [A]p)$$

This will colour every node as a win for player ***E*** first – but then, using the universal set as a first approximation, stage by stage, the right colours for ***A*** will appear. More generally, strategies seem like 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 Section 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 driving the puzzle suggest an alternative 'self-defeating' 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 needs a jump-start, viz. 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 Section 8:

B_E, B_A	–
B_A	B_E

B_E, B_A	–
	B_E

B_E, B_A	–

B_E, B_A

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 Section 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:

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

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*:

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

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

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

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

Thus a conditional is a static encoding right now of what agents would believe when updated. Beyond simple world elimination by public announcements, recent dynamic belief logics also describe how agents' plausibility relations may get *changed* as new information comes in, generalizing existing belief revision theories (Aucher 2003).

Example 14 Updating plausibilities.

Consider our running Example 1 again, but 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*.

Taken 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 resulting sequence of plausibility relations would be that players believe that they are in the solution set of the algorithm. ♣

This is just one of many ways in which dynamic epistemic analysis of game solution procedures can be refined to include beliefs, as well as other update triggers. [12]

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 (or doxastic) procedures. As in Section 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 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 & Royu 2005). At nodes of the tree, players still see a set of possible histories continuing the one so far. Further information may lead them to rule out branches from this set.

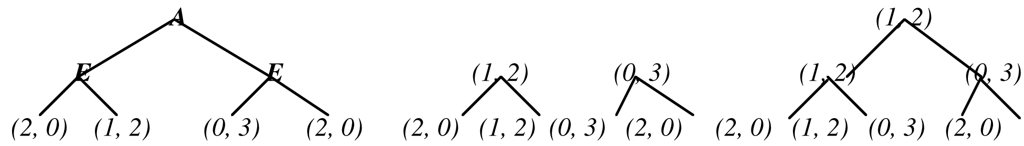
[12] A more sophisticated dynamic analysis of belief revision is found in van Benthem 2006.

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 procedure for node values. But we can also recast its steps as an elimination procedure for branches, driven by iterated announcement of an analogue of the earlier rationality principles of Section 4. Here is one version of this:

Definition 13 Momentaneous Rationality.

The assertion of *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. E.g., 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.

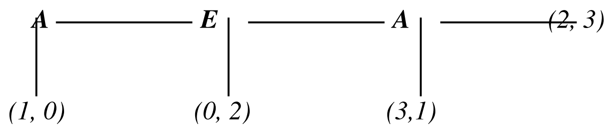
Again, this announcement scenario also suggests alternative solution procedures. For instance, a more co-operative scenario might involve an assertion of

Cooperative Rationality CR

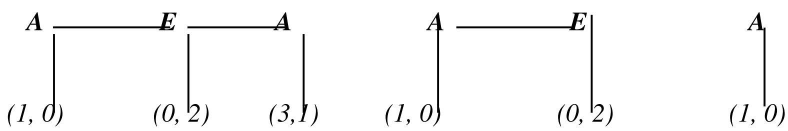
Players never select a move *m* when there is some other move allowing at least one outcome that is better for both players than any history following *m*.

Example 16 *MR* and *CR* conversation in a Centipede game.

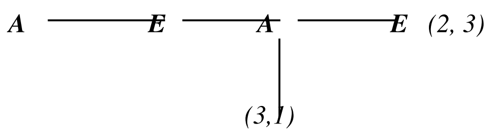
Here is a famous example, of which we just show a simple case:



Backward induction computes value *(1, 0)* for the initial node: ***A*** plays down. Now, iterated announcement of the assertion *MR* would do this in the following 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 even enforce a unique solution here, unlike *CR* by itself. One of these might be "I will repay favours", i.e., the risks of losing a certain 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. [13]

[13] At this stage, it would be tempting to now consider extensive games as they evolve over time. Players then experience two different processes, viz. *update with observed moves* plus *revision of expectations*

8 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 left implicit. Now, 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 technical findings, solving a game involves dynamic epistemic procedures which are of interest per se, and game-theoretic equilibria are then greatest fixed points of such procedures. This analogy suggests a general study of game solution concepts in dynamic epistemic logic, instead of just separate epistemic characterization theorems. Sections 5, 6 identified a number of model-theoretic results on dynamic epistemic logics which show there is content to such a connection. 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!

---

about the future course of the game. To do justice to this, we would 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 back to full-fledged strategy-profile worlds after all. Van Benthem 2002D has further discussion and a richer temporal framework for dealing with such scenarios.

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. Given all this, we certainly do not claim that bringing in 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.

9 References

K. Apt, 2005, The Many Faces of Rationalizability', CWI Amsterdam.

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

A. Baltag, L. Moss & S. Solecki, 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. Updated versions through 2004.

J. van Benthem, 1996, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

J. van Benthem, 2002A, 'Extensive Games as Process Models', *Journal of Logic, Language and Information* 11, 289–313.

J. van Benthem, 2002B, 'One is a Lonely Number: on the logic of communication', Colloquium Logicum, Muenster. Tech Report PP-2002-27, ILLC Amsterdam. To appear in *Proceedings Colloquium Logicum*, AMS Publications.

J. van Benthem, 2002C, 'Rational Dynamics and Epistemic Logic in Games', Tech Report PP-2002-26, ILLC Amsterdam.

J. van Benthem, 2002D, 'Update and Revision in Games', working paper, ILLC Amsterdam & Stanford University. Various versions through 2004.

J. van Benthem, 2005, 'Dynamic Logic of Belief Revision', ILLC Tech Report, University of Amsterdam.

J. van Benthem, J. van Eijck & B. Kooi, 2004, 'Axioms for Communication and Action', ILLC, University of Amsterdam, CWI Amsterdam & Department of Philosophy, University of Groningen.

J. van Benthem, S. van Otterloo & O. Roy, 2005, 'Preference Logic, Conditionals, and Solution Concepts in Games', manuscript, ILLC Amsterdam. To appear in *Festschrift for Krister Segerberg*, University ofUppsala.

P. Blackburn, M. de Rijke & Y. Venema, 2000, *Modal Logic*, Cambridge University Press, Cambridge.

B. de Bruin, 2003, *Explaining Games - on the Logic of Game-Theoretic Explanations*, Dissertation DS 2004-03, ILLC, University of Amsterdam.

K. Binmore, 1992, *Fun and Game*s, Heath & Company, Lexington (Mass.).

A. Dawar, E. Grädel, and S. Kreutzer, 2004, 'Inflationary Fixed Points in Modal Logic', *ACM Transations on Computational Logic*, vol. 5, pp. 282 - 315.

H. van Ditmarsch, 2002, 'Keeping Secrets with Public Communication', Department of computer science, University of Otago, Dunedin.

H. van Ditmarsch. W. van der Hoek & B. Kooi, 2004, Dynamic Epistemic Logic, to appear with Kluwer-Springer Publishers, Dordrecht.

H-D Ebbinghaus & J. Flum, 1995, *Finite Model Theory*, Springer, Berlin.

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

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

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

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

R. Stalnaker, 1999, ‘Extensive and Strategic Form: Games and Models for Games’, *Research in Economics*, 53:293-291.

C. Stirling, 1999, 'Bisimulation, Modal Logic, and Model Checking Games', *Logic Journal of the IGPL 7:1*, 103–124.