# A Dynamic Analysis of Interactive Rationality

Eric Pacuit[1] Olivier Roy[2]

[1] Tilburg Institute for Logic and Philosophy of Science, e.j.pacuit@uvt.nl
[2] Center for Mathematical Philosophy, LMU, Olivier.Roy@lrz.uni-muenchen.de

**Abstract.** Epistemic game theory has shown the importance of informational contexts in understanding strategic interaction. We propose a general framework to analyze how such contexts may arise. The idea is to view informational contexts as the fixed-points of iterated, "rational responses" to incoming information about the agents' possible choices. We show general conditions for the stabilization of such sequences of rational responses, in terms of structural properties of both the decision rule and the information update policy.

## 1 Background and Motivation

An increasingly popular[3] view is that "*the* fundamental insight of game theory [is] that a rational player must take into account that the players reason about each other in deciding how to play" [6, pg. 81]. Exactly *how* the players (should) incorporate the fact that they are interacting with other (actively reasoning) agents into their own decision making process is the subject of much debate. A variety of frameworks explicitly model the *reasoning* of rational agents in a strategic situation. Key examples include Brian Skyrms' models of "dynamic deliberation" [32], Ken Binmore's analysis of "eductive reasoning" [11], and Robin Cubitt and Robert Sugden's "common modes of reasoning" [17]. Although the details of these frameworks are quite different they share a common line of thought: In contrast to classical game theory, *solution concepts* are no longer the basic object of study. Instead, the "rational solutions" of a game are the result of individual (rational) decisions in specific informational "contexts".

This perspective on the foundations of game theory is best exemplified by the so-called epistemic program in game theory (cf. [15]). The central thesis here is that the basic mathematical model of a game should include an explicit parameter describing the players' *informational attitudes*. However, this broadly decision-theoretic stance does not simply *reduce* the question of decision-making in interaction to that of rational decision making in the face of uncertainty or ignorance. Crucially, *higher-order* information (belief about beliefs, etc.) are key components of the informational context of a game[4]. Of course, different contexts

[3] But, of course, not uncontroversial. See, for example, [22, pg. 239].

[4] That is, strategic behavior *depends*, in part, on the players' higher-order beliefs. However, the question of what precisely is being claimed should be treated with some care. The well-known *email game* of Ariel Rubinstein [30] demonstrates that

of a game can lead to drastically different outcomes, but this means that the informational contexts themselves are open to rational criticism:

> "It is important to understand that we have two forms of irrationality [...]. For us, a player is rational if he optimizes and also rules nothing out. So irrationality might mean not optimizing. But it can also mean optimizing while not considering everything possible." [16, pg. 314]

Thus, a player can be rationally criticized for not choosing what is *best given their information*, but also for not reasoning *to* a "proper" context. Of course, what counts as a "proper" context is debatable. There might be rational pressure for or against making certain *substantive assumptions*[5] about the beliefs of one's opponents, for instance, always entertaining the possibility that one of the players might not choose optimally.

Recently, researchers using methods from dynamic-epistemic logic have taken steps to understanding this idea of reasoning *to* a "proper" or "rational" context [10, 9, 8, 36]. Building on this literature[6], we provide a general characterization of when players can or cannot rationally reason to an informational context.

## 2 Belief Dynamics for Strategic Games

Our goal is to understand well-known solutions concepts, not in terms of fixed informational contexts—for instance, models (e.g., type spaces or epistemic models) satisfying rationality and common belief of rationality—but rather as a result of a dynamic, interactive process of "information exchanges". It is important to note that we do *not* see this work as an attempt to represent some type of "preplay communication" or form of "cheap talk". Instead, the idea is to represent the process of *rational deliberation* that takes the players from the *ex ante* stage to the *ex interim* stage of decision making. Thus, the "informational exchanges" are the result of the players' *practical reasoning* about what they should do, given their current beliefs. This is in line with the current research program using dynamic epistemic and doxastic logics to analyze well-known solution concepts (cf. [2, 9, 10] where the "rationality announcements" do not capture any type of communication between the players, but rather internal observations about which outcomes of the game are "rational").

misspecification of arbitrarily high-orders of beliefs can have a great impact on (predicted) strategic behavior. So there are simple examples where (predicted) strategic behavior is *too sensitive* to the players' higher-order beliefs. We are not claiming that a rational agent is *required* to consider *all* higher-order beliefs, but only that a rational player recognizes that her opponents are actively reasoning, rational agents, which means that a rational player does take into account *some* of her higher-order beliefs (e.g., what she believes her opponents believe she will do) as she deliberates. Precisely "how much" higher-order information should be taken into account is a very interesting, open question which we set aside in this paper.

[5] The notion of substantive assumption is explored in more detail in [29].

[6] The reader not familiar with this area can consult the recent textbook [35] for details.

### 2.1 Describing an Informational Context

Let $G = \langle N, \{S_i\}_{i \in N}, u_i \rangle$ be a strategic game (where $N$ is the set of players and for each $i \in N$, $S_i$ is the set of actions for player $i$ and $u_i : \Pi_i S_i \to \mathbb{R}$ is a utility function).[7] The informational context of a game describes the players' *hard* and *soft* information about the possible outcomes of the game. Many different formal models have been used to represent an informational context of a game (for a sample of the extensive literature, see [13, 10] and references therein). In this paper we employ one such model: a *plausibility structure* consisting of a set of states and a single plausibility ordering (which is reflexive, transitive and connected) $w \preceq v$ that says "$v$ is at least as plausible as $w$." Originally used as a semantics for conditionals (cf. [24]), these *plausibility models* have been extensively used by logicians [34, 35, 8], game theorists [12] and computer scientists [14, 23] to represent rational agents' (all-out) beliefs. We thus take for granted that they provide a natural model of beliefs in games:

**Definition 1.** *Let $G = \langle N, \{S_i\}_{i \in N}, u_i \rangle$ be a strategic form game. An* **informational context** *of $G$ is a plausibility model $\mathcal{M}_G = \langle W, \preceq, \sigma \rangle$ where $\preceq$ is a connected, reflexive, transitive and well-founded*[8] *relation on $W$ and $\sigma$ is a* **strategy function***: a function $\sigma : W \to \Pi_i S_i$ assigning strategy profiles to each state. To simplify notation, we write $\sigma_i(w)$ for $(\sigma(w))_i$ (similarly, write $\sigma_{-i}(w)$ for the sequence of strategies of all players except $i$).*

A few comments about this definition are in order. First of all, note that there is only one plausibility ordering in the above models, yet we are interested in games with more than one player. There are different ways to interpret the fact that there is only one plausibility ordering. One is that the models represent the beliefs of a single player before she has made up her mind about which option to choose in the game. A second interpretation is to think of a model as representing the modeler's or game theorist's point of view about which outcomes are more or less plausible given the reasoning of the players. Thus, a model describes a stage of the rational deliberation of *all* the players starting from an initial model where the players have the same beliefs (i.e., the *common prior*). The private information about which outcomes the *players* consider possible given their actual choice can then be defined from the *conditional beliefs*.[9] Our second comment on the above definition is that since we are representing the rational

[7] We assume the reader is familiar with the basic concepts of game theory. For example, strategic games and various solution concepts, such as iterated removal of strictly (weakly) dominated strategies.

[8] Well-foundedness is only needed to ensure that, for any set $X$, the set of minimal elements in $X$ is nonempty. This is important only when $W$ is infinite – and there are ways around this in current logics. Moreover, the condition of connectedness can also be lifted, but we use it here for convenience.

[9] The suggestion here is that one can define a partition model á la Aumann [5] from a plausibility model. Working out the details is left for future work, but we note that such a construction blurs the distinction between so-called *belief*-based and *knowledge*-based analyses of solution concepts (cf. the discussion in [15]).

deliberation process, we do not assume that the players have made up their minds about which actions they will choose. Finally, note that the strategy functions need not be onto. Thus, the model represents the player's(s') opinions about which outcomes of the game are more or less plausible *among the ones that have not been ruled out.*

Of course, this model can be (and has been: see [8, 35]) extended to include beliefs for each of the players, an explicit relation representing the player(s) hard information or by making the plausibility orders state-dependent. In order to keep things simple we focus on models with a single plausibility ordering.

We conclude this brief introduction to plausibility models by giving the well-known definitions of a conditional belief. For $X \subseteq W$, let $Min_{\preceq}(X) = \{v \in X \mid v \preceq w \text{ for all } w \in X\}$ be the set of minimal elements of $X$ according to $\preceq$.

**Definition 2 (Belief and Conditional Belief).** *Let $\mathcal{M}_G = \langle W, \preceq, \sigma \rangle$ be a model of a game $G$. Let $E$ and $F$ be subsets of $W$, we say:*

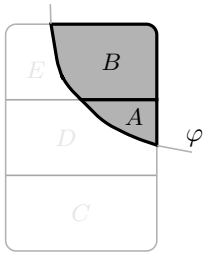
- *$E$ is* **believed conditional on** *$F$ in $\mathcal{M}_G$ provided $Min_{\preceq}(F) \subseteq E$.*

*Also, we say $E$ is* **believed** *in $\mathcal{M}_G$ if $E$ is believed conditional on $W$. Thus, $E$ is believed provided $Min_{\preceq}(W) \subseteq E$*

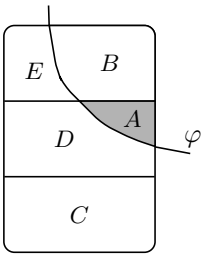
### 2.2 A Primer on Belief Dynamics

We are not interested in informational contexts *per se*, but rather how the informational context changes during the process of rational deliberation. The type of change we are interested in is how a model $\mathcal{M}_G$ of a game $G$ incorporates new information about what the players *should* do (according to a particular choice rule). As is well known from the belief revision literature, there are many ways to transform a plausibility model given some new information [28]. We do not have the space to survey the entire body of relevant literature here (cf., [35, 7]). Instead we sketch some key ideas, assuming the reader is already familiar with this approach to belief revision.

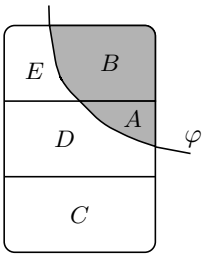
The general approach is to define a way of *transforming* a plausibility model $\mathcal{M}_G$ given a proposition $\varphi$. A transformation $\tau$ maps plausibility models and propositions to plausibility models (we write $\mathcal{M}_G^{\tau(\varphi)}$ for $\tau(\mathcal{M}_G, \varphi)$). Different definitions of $\tau$ represent the different attitudes an agent can take to the incoming information. The picture below provides three typical examples:




$!(\varphi) : A \prec B$




$\uparrow(\varphi) : A \prec C \prec D \prec B \cup E$




$\Uparrow(\varphi) : A \prec B \prec C \prec D \prec E$

The operation on the left is the well-known *public announcement* operation [25, 19], which assumes that the source of $\varphi$ is *infallible*, ruling out any possibilities that are inconsistent with $\varphi$. For the other transformations, while the players do *trust* the source of $\varphi$, they do not treat the source as infallible. Perhaps the most ubiquitous policy is *conservative upgrade* ($\uparrow\varphi$), which allows the player(s) only tentatively to accept the incoming information $\varphi$ by making the best $\varphi$-worlds the new minimal set while keeping the old plausibility ordering the same on all other worlds. The operation on the right, *radical upgrade* ($\Uparrow\varphi$), is stronger, moving *all* $\varphi$ worlds before all the $\neg\varphi$ worlds and otherwise keeping the plausibility ordering the same. These dynamic operations satisfy a number of interesting logical principles [35, 7], which we do not discuss further here.

We are interested in the operations that transform the informational context as the players deliberate about what they should do in a game situation. In each informational context (viewed as describing one stage of the deliberation process), the players determine which options are "*rationally permissible*" and which options the players ought to avoid (which is guided by some fixed choice rule). This leads to a transformation of the informational context as the players adopt the relevant beliefs about the outcome of their *practical reasoning*. The different types of transformation mentioned above then represent how confident the player(s) (or modeler) is (are) in the assessment of which outcomes are rational. In this new informational context, the players again think about what they should do, leading to another transformation. The main question is does this process *stabilize*?

The answer to this question will depend on a number of factors. The general picture is

$$\mathcal{M}_0 \overset{\tau(D_0)}{\Longrightarrow} \mathcal{M}_1 \overset{\tau(D_1)}{\Longrightarrow} \mathcal{M}_2 \overset{\tau(D_2)}{\Longrightarrow} \cdots \overset{\tau(D_n)}{\Longrightarrow} \mathcal{M}_{n+1} \Longrightarrow \cdots$$

where each $D_i$ is some proposition and $\tau$ is a model transformer. Two questions are important for the analysis of this process. First, what type of transformations are the players using? For example, if $\tau$ is a public announcement, then it is not hard to see that, for purely logical reasons, this process must eventually stop at a limit model (see [8] for a discussion and proof). The second question is where do the propositions $D_i$ come from? To see why this matters, consider the situation where you iteratively perform a radical upgrade with $p$ and $\neg p$ (i.e., $\Uparrow(p), \Uparrow(\neg p), \ldots$). Of course, this sequence of upgrades never stabilizes. However, in the context of reasoning about what to do in a game situation, this situation may not arise thanks to special properties of the choice rule that is being used to describe (or guide) the players' decisions.

### 2.3 Deliberating about What to Do

It is not our intention to have the dynamic operations of belief change discussed in the previous section directly represent the players' (practical) *reasoning*. Instead, we treat practical reasoning as a "black box" and focus on general *choice rules* that are intended to describe rational decision making (under ignorance). To make this precise, we need some notation:

**Definition 3 (Strategies in Play).** *Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game and $\mathcal{M}_G = \langle W, \preceq, \sigma\rangle$ an informational context of $G$. For each $i \in N$, the strategies* **in play for** *$i$ is the set*

$$S_{-i}(\mathcal{M}_G) = \{s_{-i} \in \Pi_{j\neq i} S_j \mid \text{ there is } w \in Min_{\preceq}(W) \text{ with } \sigma_{-i}(w) = s_{-i}\}$$

This set $S_{-i}(\mathcal{M}_G)$ is the set of strategies that are believed to be available for player $i$ at some stage of the deliberation process represented by the model $\mathcal{M}_G$. Given $S_{-i}(\mathcal{M}_G)$, different choice rules offer recommendations about which options to choose. There are many choice rules that could be analyzed here (e.g., strict dominance, weak dominance or admissibility, minimax, minmax regret, etc.). For the present purposes we focus primarily on weak dominance (or admissibility), although our main theorem in Seciton 3 applies to all choice rules.

**Weak Dominance (pure strategies[10])** Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game and $\mathcal{M}_G$ an model of $G$. For each $i$ and $a \in S_i$, put $a \in S_i^{wd}(\mathcal{M}_G)$ provided there is $b \in S_i$ such that for all $s_{-i} \in S_{-i}(\mathcal{M}_G)$, $u_i(s_{-i}, b) \geq u_i(s_{-i}, a)$ and there is some $s_{-i} \in S_{-i}(\mathcal{M}_G)$ such that $u_i(s_{-i}, b) > u_i(s_{-i}, a)$.

So an action $a$ is weakly dominated for player $i$ if it is weakly dominated with respect to all of $i$'s available actions and the (joint) strategies believed to be still in play for $i$'s opponents.

More generally, we assume that given the beliefs about which strategies are in play the players *categorize* their available options (i.e., the set $S_i$) into "good" (or "rationally permissible") strategies and those strategies that are "bad" (or "irrational"). Formally, a **categorization** for player $i$ is a pair $\mathbf{S}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$ where $S_i^+ \cup S_i^- \subseteq S_i$. (We write $\mathbf{S}_i(\mathcal{M}_G)$ to signal that the categorization depends on current beliefs about which strategies are in play.) Note that, in general, a categorization need not be a partition (i.e., $S_i^+ \cup S_i^- \neq S_i$) . See [18] for an example of such a categorization algorithm. However, in the remainder of this paper we focus on familiar choice rules where the categorization does form a partition. For example, for weak dominance we let $S_i^- = S_i^{wd}(\mathcal{M}_G)$ and $S_i^+ = S_i - S_i^-$.

Given a model of a game $\mathcal{M}_G$ and for each player $i$ a categorization is $\mathbf{S}_i(\mathcal{M}_G)$; the next step is to incorporate this information into $\mathcal{M}_G$ using some model transformation. We start by introducing a simple propositional language to describe a categorization.

**Definition 4 (Language for a Game).** *Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game. Without loss of generality, assume that each of the $S_i$ is disjoint and let $\mathsf{At}_G = \{P_a^i \mid a \in S_i\}$ be a set of atomic formulas (one for each $a \in S_i$). The propositional language for $G$, denoted $\mathcal{L}_G$, is the smallest set of formulas containing $\mathsf{At}_G$ and closed under the Boolean connectives $\neg$ and $\wedge$.*

Formulas of $\mathcal{L}_G$ are intended to describe possible outcomes of the game. Given an informational context of a game $\mathcal{M}_G$, the formulas $\varphi \in \mathcal{L}_G$ is can be associated with subsets of the set of states in the usual way:

[10] This definition can be modified to allow for dominance by mixed strategies, but we leave issues about how to incorporate probabilities to another occasion.

**Definition 5.** *Let $G$ be a strategic game, $\mathcal{M}_G = \langle W, \preceq, \sigma\rangle$ an informational context of $G$ and $\mathcal{L}_G$ a propositional language for $G$. We define a map $[\![\cdot]\!]_{\mathcal{M}_G} : \mathcal{L}_G \to \wp(W)$ by induction as follows: $[\![P_a^i]\!]_{\mathcal{M}_G} = \{w \mid \sigma(w)_i = a\}$, $[\![\neg\varphi]\!]_{\mathcal{M}_G} = W - [\![\varphi]\!]_{\mathcal{M}_G}$ and $[\![\varphi \wedge \psi]\!]_{\mathcal{M}_G} = [\![\varphi]\!]_{\mathcal{M}_G} \cap [\![\psi]\!]_{\mathcal{M}_G}$.*

Using the above language, for each informational context of a game $\mathcal{M}_G$, we can define $Do(\mathcal{M}_G)$, which describes what the players are going to do according to a fixed categorization procedure. To make this precise, suppose that $\mathbf{S}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$ is a categorization for each $i$ and define:

$$Do_i(\mathcal{M}_G) \;:=\; \bigvee_{a \in S_i^+} P_i^a \wedge \bigwedge_{b \in S_i^-} \neg P_i^b$$

Then, let $Do(\mathcal{M}_G) = \bigwedge_i Do_i(\mathcal{M}_G)$.[11]

The general project is to understand the interaction between types of categorizations (eg., choice rules) and types of model transformations (representing the rational deliberation process). One key question is: Does a deliberation process *stabilize*(and if so, under what conditions)? (See [8] for general results here.) In this paper there are two main reasons why an upgrade stream would stabilize. The first is from properties of the transformation. The second is because the choice rule satisfies a monotonicity property so that, eventually, the categorizations stabilize and no new transformations can change the plausibility ordering. We are now ready to give a formal definition of a "deliberation sequence":

**Definition 6 (Deliberation Sequence).** *Given a game $G$ and an informational context $\mathcal{M}_G$, a deliberation sequence of type $\tau$ (which we also call an upgrade sequence), induced by $\mathcal{M}_G$ is an infinite sequence of plausibility models $(\mathcal{M}_m)_{m\in\mathbb{N}}$ defined as follows:*

$$\mathcal{M}_0 = \mathcal{M}_G \qquad \mathcal{M}_{m+1} = \tau(\mathcal{M}_m, Do(\mathcal{M}_m))$$

*An upgrade sequence* **stabilizes** *if there is an $n \geq 0$ such that $\mathcal{M}_n = \mathcal{M}_{n+1}$.*

## 3 Case Study: Iterated Admissibility

A key issue in the epistemic foundations of game theory is the epistemic analysis of iterated removal of *weakly* dominated strategies. Many authors have pointed out puzzles surrounding such an analysis [4, 31, 16]. For example, Samuelson [31] showed (among other things) that "common knowledge of admissibility" may be an inconsistent concept (in the sense that there is a game which does not have a model with a state satisfying 'common knowledge of rationality' [31, Example 8, pg. 305]).[12] This is illustrated by the following game:

[11] There are other ways to describe a categorization, but we leave this for further research.

[12] Compare with strict dominance: it is well known that common knowledge that players do not play weakly dominated strategies *implies* that the players choose a strategy profile that survives iterated removal of strictly dominated strategies.

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $u$ | $1,1$ | $1,0$ |
| | $d$ | $1,0$ | $0,1$ |

The key issue is that the assumption that players only play *admissible* strategies conflicts with the logic of iteratively removing strategies deemed "irrational". The general framework introduced above offers a new, dynamic perspective on this issue, and on reasoning with admissibility more generally.[13] Dynamically, Samuelson's non-existence result corresponds to the fact that the players' rational upgrade streams do not stabilize. That is, the players are not able to deliberate their way to a stable, common belief in admissibility. In order to show this we need the "right" notion of model transformation.

Our first observation is that the model transformations we discussed in Section 2.2 do not explain Samuelson's result.

**Observation 1** Suppose that the categorization method is weak dominance and that $Do(\mathcal{M})$ is defined as above. For each of the model transformations discussed in Section 2.2 (i.e., public announcement, radical upgrade and conservative upgrade), any deliberation sequence for the above game stabilizes.

The proof of this Observation is straightforward since the language used to describe the categorization does not contain belief modalities[14]. This observation is nice, but it does not explain the phenomena noticed by Samuelson [31]. The problem lies in the way we incorporate information when there is more than one element of $S_i^+(\mathcal{M})$ for some agent $i$.

It is well known that, in general, there are no rational principles of decision making (under ignorance or uncertainty) which *always* recommend a *unique* choice. In particular, it is not hard to find a game and an informational context where there is at least one player without a *unique* "rational choice". How should a rational player incorporate the information that more than one action is classified as "choice-worthy" or "rationally permissible" (according to some choice rule) for her opponent(s)? Making use of a well-known distinction due to Edna Ullmann-Margalit and Sidney Morgenbesser [33], the assumption that all players are rational can help determine which options the player will *choose*, but rationality alone does not help determine which of the rationally permissible options will be "picked"[15]. What interests us is how to transform a plausibility

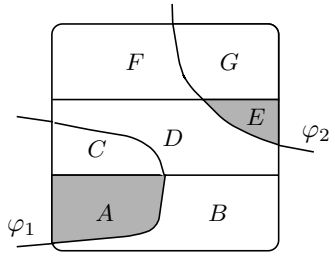
[13] We do not provide an alternative epistemic characterization of this solution concept. Both [16] and [20] have convincing results here. Our goal is to use this solution concept as an illustration of our general approach.

[14] An interesting extension would be to start with a multiagent belief model and allow players not only to incorporate information about which options are "choice-worthy", but also what beliefs their opponents may have. We leave this extension for future work, focusing here on setting up the basic framework.

[15] This line of thought led Cubitt and Sugden to impose a "privacy of tie breaking" property which says that players cannot *know* that her opponent will not pick an

model to incorporate the fact that there is a *set* of choice-worthy options for (some of) the players.

We suggest that a generalization of *conservative upgrade* is the notion we are looking for (see [21] for more on this operation). The idea is to do an upgrade with a *set* of propositions $\{\varphi_1, \ldots, \varphi_n\}$ by letting the most plausible worlds be the union of each of the most plausible $\varphi_i$ worlds:




$$\uparrow\{\varphi_1, \varphi_2\} : A \cup E \prec B \prec C \cup D \prec F \cup G$$

We do not give the formal definition here, but it should be clear from the example given above. It is not hard to see that this is not the same as $\uparrow\varphi_1 \vee \cdots \vee \varphi_n$, since, in general, $Min_{\preceq}([\![\varphi_1]\!] \cup \cdots \cup [\![\varphi_n]\!]) \neq \bigcup_i Min_{\preceq}([\![\varphi_i]\!])$. We must modify our definition of $Do(\mathcal{M})$: for each $i \in N$ let:

$$Do_i(\mathbf{S}_i(\mathcal{M}_G)) = \{P_a^i \mid a \in \mathbf{S}_i^+(\mathcal{M}_G)\} \cup \{\neg P_b^i \mid b \in \mathbf{S}_i^-(\mathcal{M}_G)\}$$

Then define $Do(\mathbf{S}(\mathcal{M}_G)) = Do_i(\mathbf{S}_i(\mathcal{M}_G)) \bigwedge Do_2(\mathbf{S}_2(\mathcal{M}_G)) \cdots \bigwedge Do_n(\mathbf{S}_n(\mathcal{M}_G))$, where if $X$ and $Y$ are two sets of propositions, then let $X \wedge Y := \{\varphi \wedge \psi \mid \varphi \in X, \psi \in Y\}$.

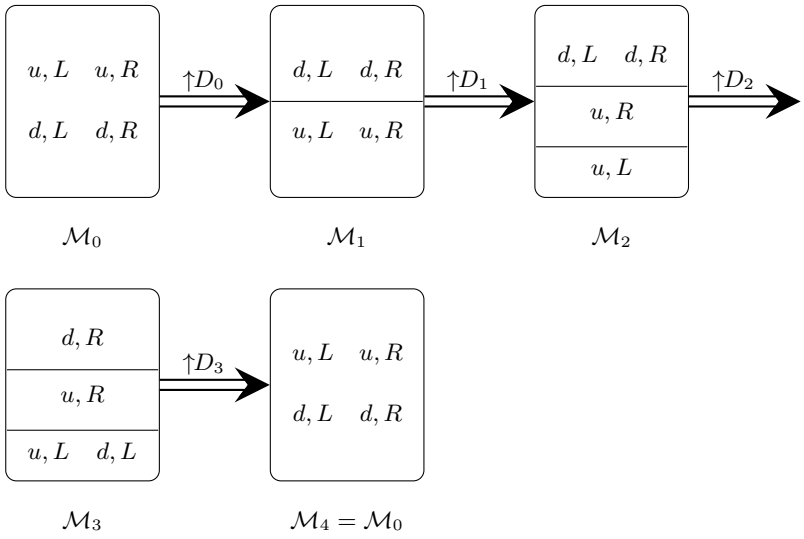
**Observation 2** Suppose that the categorization method is weak dominance as explained in Section 2.3 and that $Do(\mathcal{M})$ is defined as above. Then, starting with the initial full model of the above game,[16] a generalized conservative upgrade stream does not stabilize.

The following upgrade stream illustrates this observation:

---

option that is classified as "choice-worthy" [17, pg. 8] (cf. also [4]'s "no extraneous restrictions on beliefs" property). Wlodeck Rabinovich takes this even further and argues that from the principle of indifference, players must assign equal probability to all choice-worthy options [27].

[16] A full model is one where it is common knowledge that each outcome of the game is equally plausible.




Intuitively, from $\mathcal{M}_0$ to $\mathcal{M}_2$ the agents have reasons to exclude $d$ and $R$, leading them to the common belief that $u, L$ is played. At that stage, however, $d$ is admissible for Ann, canceling the reason the agents had to rule out this strategy. The rational response here is thus to suspend judgment on $d$, leading to $\mathcal{M}_3$. In this new model the agents are similarly led to suspend judgment on not playing $R$, bringing them back to $M_0$. This process loops forever: the agents' reasoning does not stabilize.

A corollary of this observation is that common belief in admissibility is not sufficient for the stabilization of upgrade streams. Stabilization also requires that all *and only* those profiles that are most plausible are admissible.

## 4 Stabilization Theorem

In this section we informally state and discuss a number of abstract principles which guarantee that a rational deliberation sequence will *stabilize*. The principles ensure that the categorizations are "sensitive" to the players' beliefs and that the players respond to the categorizations in the appropriate way.

We start by fixing some notation. Let $U$ be a fixed set of states and $G$ a fixed strategic game. We confine our attention to transformations between models of $G$ whose states come from the universe of states $U$. Let $\mathbb{M}_G$ be the set of all such plausibility models. A model transformation is then a function that maps a model of $G$ and a finite set of formulas of $\mathcal{L}_G$ to a model in $\mathbb{M}_G$:

$$\tau : \mathbb{M}_G \times \wp_{<\omega}(\mathcal{L}_G) \to \mathbb{M}_G$$

where $\wp_{<\omega}(\mathcal{L}_G)$ is the set of finite subsets of $\mathcal{L}_G$. Of course, not all transformations $\tau$ make sense in this context.

The first set of principles that $\tau$ must satisfy ensure that the categorizations and belief transformation $\tau$ are connected in the "right way". One natural property is that the belief transformations treat *equivalent* formulas the same way. A second property we impose is that receiving exactly the same (ground) information twice does not have any effect on the players' beliefs. These are general properties of the belief transformation. Certainly, there are other natural properties that one may want to impose (for example, variants of the AGM postulates [1]), but for now we are interested in the minimal principles needed to prove a stabilization result.

The next set of properties ensure that the transformations respond "properly" to a categorization. First, we need a property to guarantee that the categorizations depend only on the players' beliefs. Second, we need to ensure that all upgrade sequences respond to the categorizations in the right way:

**C2**$^-$ For any upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$ in $\tau$, if $a \in S_i^-(\mathcal{M}_n)$ then $\neg P_i^a$ is believed in $\mathcal{M}_{n+1}$.

**C2**$^+$ For any upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$ in $\tau$, if $a \in S_i^+(\mathcal{M}_n)$ then $\neg P_i^a$ is not believed in $\mathcal{M}_{n+1}$

Finally, we need to assume that the categorizations are monotonic:

**Mon**$^-$ For any upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$, for all $n \geq 0$, for all players $i \in N$, $\mathbf{S}_i^-(\mathcal{M}_n) \subseteq \mathbf{S}_i^-(\mathcal{M}_{n+1})$

**Mon**$^+$ Either for all models $\mathcal{M}_G$, $\mathbf{S}_i^+(\mathcal{M}_G) = S_i - \mathbf{S}_i^-(\mathcal{M}_G)$ or for any upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$, for all $n \geq 0$, for all players $i \in N$, $\mathbf{S}_i^+(\mathcal{M}_n) \subseteq \mathbf{S}_i^+(\mathcal{M}_{n+1})$

In particular, **Mon**$^-$ means that once an option for a player is classified as "not rationally permissible", it cannot drop this classification at a later stage of the deliberation process.

**Theorem 3.** *Suppose that $G$ is a finite game and all of the above properties are satisfied. Then every upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$ stabilizes.*

The proof can be found in the full version of the paper. The role of monotonicity of the choice has been noticed by a number of researchers (see [3] for a discussion). This theorem generalizes van Benthem's analysis of rational dynamics [10] to soft information, both in terms of attitudes and announcements. It is also closely related to the result in [3] (a complete discussion can be found in the full paper).

## 5 Concluding remarks

In this paper we have proposed a general framework to analyze how "proper" informational contexts my arise. We have provided general conditions for the stabilization of deliberation sequences in terms of structural properties of both the

decision rule and the information update policy. We have also applied the framework to admissibility, giving a dynamic analysis of Samuelson's non-existence result.

Throughout the paper we have worked with (logical) models of *all out* attitudes, leaving aside probabilistic andgraded beliefs, even though the latter are arguably most widely used in the current literature on epistemic foundations of game theory. It is an important but non-trivial task to transpose the dynamic perspective on informational contexts that we advocate here to such probabilistic models. This we leave for future work.

Finally, we stress that the dynamic perspective on informational contexts is a natural complement and not an alternative to existing epistemic characterizations of solution concepts [37], which offer rich insights into the consequences of taking seriously the informational contexts of strategic interaction. What we have proposed here is a first step towards understanding how or why such contexts might arise.

## References


1. Alchourrón, C. E., Gärdenfors, P., and Makinson, D. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic 50* (1985), 510 – 530.
2. Apt, K., and Zvesper, J. Public announcements in strategic games with arbitrary strategy sets. In *Proceedings of LOFT 2010* (2010).
3. Apt, K., and Zvesper, J. The role of monotonicity in the epistemic analysis of strategic games. *Games 1*, 4 (2010), 381 – 394.
4. Asheim, G., and Dufwenberg, M. Admissibility and common belief. *Game and Economic Behavior 42* (2003), 208 – 234.
5. Aumann, R. Interactive epistemology I: Knowledge. *International Journal of Game Theory 28* (1999), 263–300.
6. Aumann, R., and Dreze, J. Rational expectations in games. *American Economic Review 98* (2008), 72 – 86.
7. Baltag, A., and Smets, S. ESSLLI 2009 course: Dynamic logics for interactive belief revision. Slides available online at `http://alexandru.tiddlyspot.com/#%5B%5BESSLLI09%20COURSE%5D%5D`, 2009.
8. Baltag, A., and Smets, S. Group belief dynamics under iterated revision: Fixed points and cycles of joint upgrades. In *Proceedings of Theoretical Aspects of Rationality and Knowledge* (2009).
9. Baltag, A., Smets, S., and Zvesper, J. Keep 'hoping' for rationality: a solution to the backwards induction paradox. *Synthese 169* (2009), 301–333.
10. Benthem, J. v. Rational dynamics and epistemic logic in games. *International Game Theory Review 9*, 1 (2007), 13–45.
11. Binmore, K. Modeling rational players: Part I. *Economics and Philosophy 3* (1987), 179 – 214.
12. Board, O. Dynamic interactive epistemology. *Games and Economic Behavior 49* (2004), 49 – 80.
13. Bonanno, G., and Battigalli, P. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics 53*, 2 (June 1999), 149–225.

14. Boutilier, C. *Conditional Logics for Default Reasoning and Belief Revision.* PhD thesis, University of Toronto, 1992.
15. Brandenburger, A. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory 35* (2007), 465–492.
16. Brandenburger, A., Friedenberg, A., and Keisler, H. J. Admissibility in games. *Econometrica 76* (2008), 307–352.
17. Cubitt, R., and Sugden, R. Common reasoning in games: A Lewisian analysis of common knowledge of rationality. CeDEx Discussion Paper, 2011.
18. Cubitt, R., and Sugden, R. The reasoning-based expected utility procedure. *Games and Economic Behavior* (2011), In Press.
19. Gerbrandy, J. *Bisimulations on Planet Kripke.* PhD thesis, University of Amsterdam, 1999.
20. Halpern, J., and Pass, R. A logical characterization of iterated admissibility. In *Proceedings of the Twelfth Conference on Theoretical Aspects of Rationality and Knoweldge* (2009), A. Heifetz, Ed., pp. 146 – 155.
21. Holliday, W. Trust and the dynamics of testimony. In *Logic and Interaction Rationality: Seminar's Yearbook 2009* (2009), ILLC Technical Reports, pp. 147 – 178.
22. Kadane, J. B., and Larkey, P. D. Subjective probability and the theory of games. *Management Science 28*, 2 (1982), 113–120.
23. Lamarre, P., and Shoham, Y. Knowledge, certainty, belief and conditionalisation. In *Proceedings of the International Conference on Knoweldge Representation and Reasoning* (1994), pp. 415 – 424.
24. Lewis, D. *Counterfactuals.* Blackwell Publishers, Oxford, 1973.
25. Plaza, J. Logics of public communications. In *Proceedings, 4th International Symposium on Methodologies for Intelligent Systems* (1989), M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. Ras, Eds., pp. 201–216 (republished as [26]).
26. Plaza, J. Logics of public communications. *Synthese: Knowledge, Rationality, and Action 158*, 2 (2007), 165 – 179.
27. Rabinowicz, W. Tortous labyrinth: Noncooperative normal-form games between hyperrational players. In *Knowledge, Belief and Strategic Interaction* (1992), C. Bicchieri and M. L. D. Chiara, Eds., pp. 107 – 125.
28. Rott, H. Shifting priorities: Simple representations for 27 itereated theory change operators. In *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg* (2006), H. Lagerlund, S. Lindström, and R. Sliwinski, Eds., vol. 53 of *Uppsala Philosophical Studies*, pp. 359 – 384.
29. Roy, O., and Pacuit, E. Substantive assumptions and the existence of universal knowledge structures: A logical perspective. Under submission, 2010.
30. Rubinstein, A. The electronic mail game: A game with almost common knowledge. *American Economic Review 79* (1989), 385 – 391.
31. Samuelson, L. Dominated strategies and common knowledge. *Game and Economic Behavior 4* (1992), 284 – 313.
32. Skyrms, B. *The Dynamics of Raitonal Deliberation.* Harvard University Press, 1990.
33. Ullmann-Margalit, E., and Morgenbesser, S. Picking and choosing. *Social Research 44* (1977), 757 – 785.
34. van Benthem, J. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics 14*, 2 (2004), 129 – 155.
35. van Benthem, J. *Logical Dynamics of Information and Interaction.* Cambridge University Press, 2010.

36. van Benthem, J., and Gheerbrant, A. Game solution, epistemic dynamics and fixed-point logics. *Fund. Inform. 100* (2010), 1–23.
37. van Benthem, J., Pacuit, E., and Roy, O. Towards a theory of play: A logical perspective on games and interaction. *Games 2*, 1 (2011), 52–86.