Chapter 6
A Dynamic Analysis of Interactive Rationality

Eric Pacuit and Olivier Roy

Abstract Epistemic game theory has shown the importance of informational contexts to understand 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 discuss conditions under which such fixed points may exist. In the process, we generalize existing rules for information updates used in the dynamic epistemic logic literature. We then apply this framework to weak dominance. Our analysis provides a new perspective on a well known problem with the epistemic characterization of iterated removal of weakly dominated strategies.

Keywords Game theory • Dynamic epistemic logic • Rationality • Update • Fixed points • Admissibility

6.1 Introduction

A crucial assumption underlying classical game-theoretic analyses is that there is *common knowledge* that all the players are *rational*. Rationality, here, is understood in the decision-theoretic sense: The players' choices are *optimal* according to some choice rule (such as maximizing subjective expected utility). Recent work in *epistemic game theory* has focused on developing sophisticated mathematical models to study the implications of assuming that all the players are rational and that this is commonly known (or commonly believed).[1] However, if common

[1]See Perea (2012), Dekel and Siniscalchi (2015), and Pacuit and Roy (2015) for surveys of this literature.

E. Pacuit (✉)
Department of Philosophy, University of Maryland, College Park, MD, USA
e-mail: epacuit@umd.edu

O. Roy
Institut für Philosophie, Universität Bayreuth, Bayreuth, Germany
e-mail: olivier.roy@uni-bayreuth.de

J. Redmond et al. (eds.), *Epistemology, Knowledge and the Impact of Interaction*, Logic, Epistemology, and the Unity of Science 38, DOI 10.1007/978-3-319-26506-3_6

knowledge of rationality is to have an "explanatory" role in the analysis of a game-theoretic situation, then it is not enough to simply *assume* that it has obtained in an informational context of a game. It is also important to describe how the players were able to arrive at this crucial state of information.[2]

There is a growing body of literature focused on analyzing games in terms of the "process of deliberation" that leads the players to select their component of a rational outcome. Many different frameworks have been used to represent this process of deliberation:

1. John Harsanyi's "tracing procedure" identifies a unique Nash equilibrium in any finite strategic game that is the limit of a sequence of Nash equilibria from a related set of strategic games. Harsanyi thought of the tracing procedure as "a mathematical formalization of the process by which rational players coordinate their choices of strategies" (Harsanyi 1975).
2. Brian Skyrms assumes that players deliberate by calculating their subjective expected utility and then use the results of their calculations to adjust their probabilities about what they are going to do and what they expect their opponents to do (Skyrms 1990).
3. Robin Cubitt and Robert Sugden apply David Lewis's "common modes of reasoning" to game-theoretic situations. They describe the players' process of deliberation as an iterative procedure for classifying strategies (Cubitt and Sugden 2011, 2014).
4. Johan van Benthem and colleagues use ideas from dynamic epistemic logic to characterize solution concepts as fixed points of iterated "(virtual) rationality announcements" (Baltag et al. 2009; van Benthem 2014).

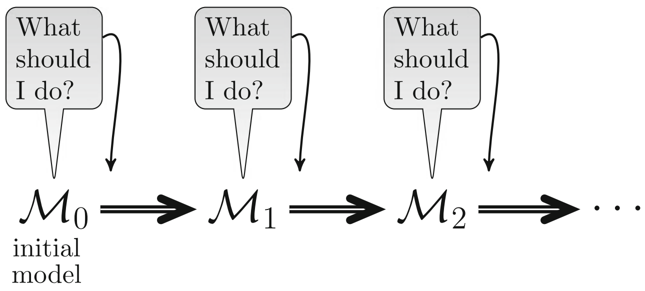
Although the details of these frameworks[3] are different, they share a common line of thought: The rational outcomes of a game are arrived at through a process in which each player settles on an optimal choice given her evolving beliefs about her own and her opponents' choices. This is not intended to be a formal account of the players' *practical reasoning* in game situations. Rather, the goal is to describe deliberation in terms of a sequence of belief changes about what the players are doing and what their opponents may be thinking. The general conclusion is that the rational outcomes of a game depend not only on the structure of the game and the players' initial beliefs, but also on which dynamical rule the players are using to update their inclinations and beliefs, and what exactly is commonly known about the process of deliberation. For instance, the outcomes of Harsanyi's tracing procedure and Skryms's model of dynamic deliberation are qualitatively similar: Both procedures lead players to choose their component of a Nash equilibrium. However, in Skyrms's model, the rate of convergence depends on the players'

[2]David Lewis already appreciated this general point about common knowledge when he first formulated his notion of common knowledge (Lewis 1969). See Cubitt and Sugden (2003) for an illuminating discussion and a reconstruction of Lewis' notion of common knowledge, with applications to game theory.

[3]See Pacuit (2015) for an extensive discussion of these different models of deliberation in games.

initial beliefs and the dynamical rule changing the players' inclinations during deliberation; and this, in turn, suggests a more refined analysis of Nash equilibrium (Skyrms 1990, pp. 154–158).

There are two key components of the above models of deliberation in games. The first component is a formal representation of the players' *state of indecision*. This is intended to be a "snapshot" of the players' *inclinations* about what they are going to choose and their *beliefs* about their opponents' choices and beliefs during the process of deliberation. The second component is the dynamical rule that governs the changes in the players' state of indecision. The general idea is that, at each stage of the deliberation, the players determine which of their available strategies are "optimal" and which they ought to avoid. Typically, it is assumed that the players are guided by some decision-theoretic choice rule, such as maximizing expected utility or avoiding dominated strategies. Using the information about the players' own choices and what they expect their opponents to do, the players' state of indecision is *transformed* according to some fixed dynamical rule. The picture to keep in mind is:



Deliberation concludes when the players reach a fixed point in the above process. The central question is: What types of transformations match different game-theoretic analyses?

In this paper, we develop a model of deliberation and characterize whether players will reason to specific informational contexts (Sect. 6.2). We then apply this framework to issues surrounding the epistemic characterization of *iterated elimination of weakly dominated strategies* (IEWDS), aka *iterated admissibility* (Sect. 6.3). Our approach builds on earlier work that describes deliberation in games in terms of (virtual) rationality announcements (van Benthem 2007; Baltag et al. 2009; Baltag and Smets 2009; van Benthem and Gheerbrant 2010).

6.2 Belief Dynamics for Strategic Games

The main idea of this paper is to understand well-known solution 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 deliberation process. It is important to note that the goal is *not* to represent some type of "pre-play communication" or some form of "cheap talk". Instead, the goal 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. In this section, we introduce our framework, incorporating ideas from the extensive literature on dynamic logics of belief revision (van Benthem 2010; Baltag and Smets 2009) and recent work on the reasoning-based approach to game theory found in Cubitt and Sugden (2011, 2014).

6.2.1 Strategic Games and Game Models

A finite strategic game is a tuple $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$, where N is a finite set of players; for each $i \in N$, S_i is a finite set of actions (also called strategies) for player i; and for each $i \in N$, $u_i : \Pi_{i\in N} S_i \to \mathbb{R}$ is a utility function assigning real numbers to each outcome of the game.[4] A **strategy profile** is a tuple $\vec{s} \in \Pi_{i\in N} S_i$, specifying an action for each player. Following standard game-theoretic notation, we write $\vec{s}_{-i} \in \Pi_{j\in N-\{i\}} S_i$ for a sequence of actions for all players except i. For simplicity, we assume that the outcomes of the game G are identified with the set of strategy profiles $S = \Pi_{i\in N} S_i$.

A **game model** describes the players' *hard* and *soft* information about the possible outcomes of the game. The models that we use in this paper are standard in the belief revision literature: a non-empty set of states, where each state is associated with a possible outcome of the game, and a single relation $\preceq$ on W representing the players' (common) initial plausibility ordering. Originally used as a semantics for conditionals (cf. Lewis 1973), these *plausibility models* have been extensively used by logicians (van Benthem 2004, 2010; Baltag and Smets 2009), game theorists (Board 2004) and computer scientists (Boutilier 1992; Lamarre and Shoham 1994) to represent rational agents' (all-out) beliefs. Thus, we take for granted that they provide natural models of (multiagent) beliefs and focus on how they can be used to represent "rational deliberation" in a game situation. The formal definition of a game model is as follows.

Definition 6.2.1 (Strategy Functions). Suppose that W is a non-empty set of states, and $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ is a finite strategic game. A **strategy function** on W for G is a function $\sigma : W \longrightarrow S$ 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).

[4] We assume that the reader is familiar with the basic concepts of game theory (e.g., strategic games and various solution concepts such as iterated removal of strictly/weakly dominated strategies). Consult Leyton-Brown and Shoham (2008) for an introduction to game theory.

Definition 6.2.2 (Game Model). Suppose that $G = \langle N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N} \rangle$ is a finite strategic game. A **model** of G is a tuple $\mathcal{M}_G = \langle W, \preceq, \sigma \rangle$, where W is a non-empty set; $\preceq$ is a connected, reflexive, transitive and well-founded[5] relation on W; and σ is a strategy function on W for G. Subsets of W are called **events**, or **propositions**.

Note that there is only one plausibility ordering in the above model; 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. First, we can take the perspective of a single player thinking about what she is going to choose in the game. Alternatively, we can think of the model as describing a stage of the rational deliberation of *all* the players, starting from a situation in which the players have the same beliefs (i.e., there is a *common prior*). The players' private beliefs, *given their actual choice of strategy*, can be defined using conditional beliefs.[6] We first need some notation. For $\emptyset \neq X \subseteq W$, 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 $\preceq$. This set contains the *most plausible* states in X.

Definition 6.2.3 (Belief and Conditional Belief). Suppose that $\mathcal{M}_G = \langle W, \preceq, \sigma \rangle$ is a model of a finite strategic game G. For all subsets E and F of W, E is believed conditional on F is defined as follows:

$$B(E \mid F) = \{w \mid Min_{\preceq}(F) \subseteq E\}.$$

We also write $B^F(E)$ for $B(E \mid F)$. If $w \in B^F(E)$, then we say that "E is believed conditional on F at w". Also, we say that E is **believed** in $\mathcal{M}_G$ if E is believed conditional on W. Thus, E is believed when $Min_{\preceq}(W) \subseteq E$.

Of course, the game models from Definition 6.2.2 can be (and have been: see Baltag and Smets 2009; van Benthem 2010) be extended to include plausibility orderings for each player, state-dependent plausibility ordering(s), explicit relations representing the players' knowledge about the game situation, and other notions of beliefs (e.g., *strong belief* or *robust belief*). To keep things simple, we focus on models with a single plausibility ordering.

[5]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.

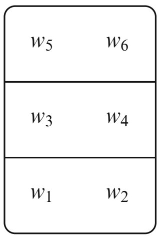
[6]A similar idea is found in standard models of differential information from the economics literature. In such models, it is assumed that there is a prior probability measure describing the players' initial beliefs (often it is the same probability measure for all the players). The players' *posterior probabilities* are defined by conditioning their prior probability measure on their private information (typically represented by some partition over the set of states).

6.2.2 A Primer on Belief Dynamics

We are not interested in game models per se, but, rather, how a game model changes during the process of rational deliberation. The type of changes 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 some decision-theoretic choice rule). As is well known from the belief revision literature, there are many ways to transform a plausibility model given some new information (Rott 2006). We do not have the space to survey this entire literature here (see van Benthem (2010) and Pacuit (2013) for modern introductions). Instead, we sketch some key ideas.

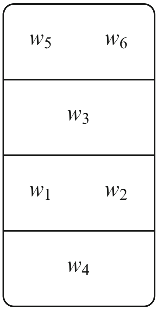
The general approach is to define a way of *transforming* a game model $\mathcal{M}_G$ given an event E. That is, we will define functions τ that map game models and events to game models. For each game model $\mathcal{M}_G$ and event E, we write $\mathcal{M}_G^{\tau(E)}$ for $\tau(\mathcal{M}_G, E)$. So, given a model $\mathcal{M}_G$ of a game G and an event E describing what the players (might/should/will) do, $\mathcal{M}_G^{\tau(E)}$ is the updated game model, taking this information into account. Different definitions of τ represent the different attitudes that an agent can have towards the incoming information.

We start with an illustrative example. Suppose that $\mathcal{M}_G = \langle W, \preceq, \sigma\rangle$ is a game model in which $W = \{w_1, w_2, w_3, w_4, w_5, w_6\}$, and $\preceq$ is defined as follows: $w_1 \sim w_2 \prec w_3 \sim w_4 \prec w_5 \sim w_6$, where $w \prec v$ means $w \preceq v$ and $v \not\preceq w$ and $w \sim v$ means $w \preceq v$ and $v \preceq w$. This game model is pictured as follows:

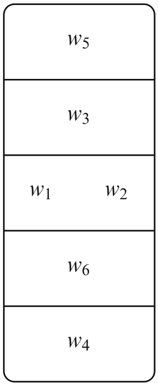


The first transformation that we discuss is the well-known *public announcement* operation (Plaza 1989; Gerbrandy 1999), denoted by !. This operation assumes that the players considers the source of the new information E *infallible*, ruling out any states not contained in E. That is, the updated model $\mathcal{M}_G^{!(E)}$ is $\langle E, \preceq', \sigma'\rangle$, where $\preceq' = \preceq \cap\, E$ and σ' is σ restricted to E.

Two other transformations have been widely discussed in the belief revision literature. For these transformations, the players do *trust* the source of the new information, though they do not treat the source as infallible. Perhaps the most ubiquitous transformation is *conservative upgrade* ($\uparrow(E)$), which lets the players only tentatively accept the incoming information E by making the most plausible E-states the new minimal set and keeping the old plausibility ordering the same on all other states. A second transformation is *radical upgrade* ($\Uparrow(E)$), which moves *all* the states in E below all the other states and, otherwise, keeps the plausibility ordering the same. The results of these two operations with $E = \{w_4, w_6\}$ on the above model $\mathcal{M}_G$ are:



$\mathcal{M}_G^{\uparrow(E)}$



$\mathcal{M}_G^{\Uparrow(E)}$

These transformations satisfy a number of interesting logical principles (van Benthem 2010) that we do not discuss in this paper.

We are interested in using these transformations to model the players' process of deliberation in a game. Given a game model (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 (as specified by some decision-theoretic choice rule). Given this information, the players use one of the above transformations to change the game model. In this new game model, the players reconsider 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 event and τ is a model transformer (e.g., public announcement, radical upgrade or conservative upgrade). Two questions are important for the analysis of this process. First, what type of transformations are the players using? For example, if τ is the public announcement transformation, then it is not hard to see that, for purely logical reasons, this process must eventually stop at a limit model (see Baltag and Smets (2009) for a discussion and proof). Second, where do the propositions D_i come from? To see why this matters, consider the situation in which you iteratively perform a radical upgrade with E and $\overline{E}$ (the complement of

E). 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 because of special properties of the choice rule that is being used to generate the events D_i.

6.2.3 Categorizing Strategies

Any sequence of game models can be viewed as the stages of a process of deliberation in the underlying game. We are interested primarily in sequences of game models that are generated by some fixed belief transformation (such as a public announcement, a conservative upgrade, or a radical upgrade). However, it is not enough to simply fix an initial game model and some model transformation to represent the players' deliberation about what they are going to do in a game. We also need a way to define the events used to update the models at each stage of the deliberation. These events should specify, for each player, which actions are "rationally permissible" and which actions they should avoid. In this section, we discuss the key features of any general method that can be used to identify the events that will serve as input to the model transformation at each stage of the deliberation.

We start with two general observations about decision making in games to motivate the definitions in this section. The first observation is that, in general, there are no rational principles of "rational" decision making (under ignorance or uncertainty) that *always* recommend a *unique* choice.[7] In particular, it is not hard to find a game and a game model where there is at least one player without a *unique* "rational choice". Making use of a well-known distinction of Ullmann-Margalit and Morgenbesser (1977), the assumption that all players are rational can help determine which options the player ought to *choose*. However, since nothing distinguishes between these on rationality grounds alone, the player is left to *pick* any of the rationally permissible options.[8]

The second observation is that we do not intend our model of deliberation to directly represent the *practical reasoning* leading to the players' decision about what to do in a game situation. In fact, we do not directly represent any formal model of practical reasoning. Instead, we treat practical reasoning as a "black box" and focus on general *choice rules* that are intended to describe the outcome of the players' practical reasoning. More generally, following Cubitt and Sugden (2014), we assume that during each stage of deliberation, the players can *categorize* their available actions. To make this precise, we need some notation:

[7] Consult any textbook on decision theory, such as Peterson (2009), for evidence of this fact.

[8] See Roy et al. (2014) and Anglberger et al. (2015) for a discussion on the rational obligations and permissions in games.

Definition 6.2.4 (Strategies in Play). Suppose that $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ is a finite strategic game and $\mathcal{M}_G = \langle W, \preceq, \sigma\rangle$ is a model of G. For each $i \in N$, the strategies **in play for** i is the set

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

The set $S_{-i}(\mathcal{M}_G)$ contains the strategies that player i still *believes* are possible at some stage of the deliberation process represented by the model $\mathcal{M}_G$. Given these beliefs, we assume that each player can *categorize* her available options:

Definition 6.2.5 (Categorization). Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game and $\mathcal{M}_G = \langle W, \preceq, \sigma\rangle$ a model of G. A **categorization** for player i in $\mathcal{M}_G$ is a pair $\mathbf{S}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$ where $S_i^+ \cup S_i^- \subseteq S_i$, $S_i^+ \cap S_i^- = \emptyset$, and

$$(*) \qquad \text{for each } a \in S_i, \text{if there is no } v \in W \text{ with } \sigma_i(v) = a, \text{then } a \in S_i^-.$$

If $\mathbf{S}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$, we write $\mathbf{S}_i^+(\mathcal{M}_G)$ for S_i^+ and $\mathbf{S}_i^-(\mathcal{M}_G)$ for S_i^-. Also, we write $\mathbf{S}(\mathcal{M}_G)$ for the sequence of categorizations $(\mathbf{S}_i(\mathcal{M}_G))_{i\in N}$.

The intended interpretation is that player i ought to pick from among the strategies in $\mathbf{S}_i^+(\mathcal{M}_G)$ and ought to avoid any strategy in $\mathbf{S}_i^-(\mathcal{M}_G)$. The strategies in $S_i - (\mathbf{S}_i^+(\mathcal{M}_G) - \mathbf{S}_i^-(\mathcal{M}_G))$ have not yet been categorized. These are the strategies that player i needs to think more about before categorizing. Condition $(*)$ in the above definition ensures that players will not choose any strategy that has been completely ruled out. Note that, in general, a categorization need not be a partition of player i's strategies (i.e., $\mathbf{S}_i^+(\mathcal{M}_G) \cup \mathbf{S}_i^-(\mathcal{M}_G) \neq S - i$). See Cubitt and Sugden (2011) for an example of such a categorization. However, many of the familiar choice rules found in the game theory literature lead to categorizations that do form a partition. Two standard examples are weak and strong dominance: Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a strategic game and $\mathcal{M}_G$ a model of G. Then:

Strong Dominance (pure strategies): For each $i \in N$, $\mathbf{SD}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$ is defined as follows: For all $a \in S_i$,

$$a \in S_i^- \text{ iff there is } b \in S_i \text{ such that for all } s_{-i} \in S_{-i}(\mathcal{M}_G),\ u_i(s_{-i}, b) > u_i(s_{-i}, a),$$

and $S_i^+ = S_i - S_i^-$.

Weak Dominance (pure strategies): For each $i \in N$, $\mathbf{WD}_i(\mathcal{M}_G) = (S_i^+, S_i^-)$ is defined as follows: For all $a \in S_i$,

$$a \in S_i^- \text{iff there is } b \in S_i \text{such that for all } s_{-i} \in S_{-i}(\mathcal{M}_G),\ u_i(s_{-i}, b) \geq u_i(s_{-i}, a)$$

and there is some

$$s_{-i} \in S_{-i}(\mathcal{M}_G) \text{ such that } u_i(s_{-i}, b) > u_i(s_{-i}, a),$$

and $S_i^+ = S_i - S_i^-$.

Both of the above definitions can be modified to cover strict/weak dominance by *mixed strategies*, but we leave issues about how to incorporate probabilities into the framework sketched in this paper for another time.

We conclude this section by defining when a game model *incorporates* a categorization. Suppose that $\mathcal{M}_G = \langle W, \preceq, \sigma \rangle$ is a game model for G, and $\mathbf{S}(\mathcal{M}'_G)$ is a categorization for a game model $\mathcal{M}'_G$. We say that $\mathcal{M}_G$ **incorporates** $\mathbf{S}(\mathcal{M}'_G)$ provided that for all $i \in N$:

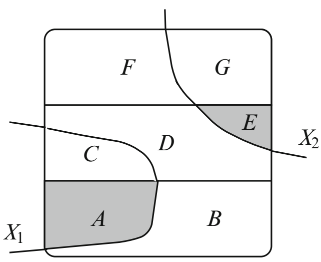
- If $a \in \mathbf{S}_i^+(\mathcal{M}'_G)$, then there is some $w \in Min_{\preceq}(W)$ such that $\sigma_i(w) = a$.
- If $a \in \mathbf{S}_i^-(\mathcal{M}'_G)$, then there is no $w \in Min_{\preceq}(W)$ such that $\sigma_i(w) = a$.

Thus, a model $\mathcal{M}_G$ incorporating a categorization $(S_i^+, S_i^-)_{i \in N}$ implies that (1) for each $a \in S_i^+$, the players do not believe that i will not play a; and (2) for each $a \in S_i^-$, players believe that i will not play a.

6.2.4 Generalized Belief Transformations

An important feature of a categorization is that more than one strategy may be "rationally permissible" for a player. This means that the information the players gain from a categorization should be represented by a *set* of events rather than a single event. Each event in this set describes the outcomes of the game that result from assuming that each player picks a rationally permissible strategy. In this section, we show how to generalize the model transformations introduced in Sect. 6.2.2 to accept finite sets of events as inputs.

Suppose that $\{E_1, \ldots, E_k\}$ is a set of events for game model $\mathcal{M}_G$. The generalization of the public announcement transformation is straightforward: $!\{E_1, \ldots, E_k\}) = !(E_1 \cup E_2 \cup \cdots E_k)$. The generalizations of the conservative and radical upgrade is more subtle. To see the difficulty, consider the game model pictured below with two events, X_1 and X_2:



The sets A, B, C, D, E, F and G denote all the different subsets of states (so, $W = A \cup B \cup C \cup D \cup E \cup F \cup G$). The plausibility ordering runs from the top to the bottom. So, for instance, the states in $A \cup B$ are the most plausible overall, and all states within $A \cup B$ are equiplausible. A conservative upgrade with $X_1 \cup X_2$ results in the following modification of the above plausibility ordering:

$$A \prec B \prec C \cup D \cup E \prec F \cup G,$$

where for sets X and Y, we write $X \prec Y$ when for all $w \in X, v \in Y, w \prec v$. However, suppose that X_1 and X_2 describe two different outcomes of the game. Furthermore, in each of the outcomes, assume that the players pick a rationally permissible action. The result of the radical upgrade with $X_1 \cup X_2$ is that the players come to believe that the outcome of the game will be as described by X_1. The beliefs are the same after a radical upgrade with $X_1 \cup X_2$, though the resulting plausibility ordering is different.

However, if both X_1 and X_2 describe situations in which all the players choose *rationally*, then why should the players believe that outcomes in X_1 are more plausible than outcomes in X_2? Researchers interested in the epistemic foundations of iterated removal of weakly dominated strategies have discussed this issue (Cubitt and Sugden 2003; Samuelson 1992). For instance, Cubitt and Sugden impose a "privacy of tie-breaking" property, which says that a player cannot *know* that her opponent will not pick an option that is classified as "choice-worthy" (Cubitt and Sugden 2014, p. 8).[9] In our setting, this issue arises because, in general, for events $E_1, \ldots, E_k$:

$$Min_{\preceq}(E_1 \cup E_2 \cup \cdots \cup E_k) \neq Min_{\preceq}(E_1) \cup Min_{\preceq}(E_2) \cup \cdots \cup Min_{\preceq}(E_k).$$

Returning to our example in the previous paragraph, the gray shaded regions identify the most plausible states in X_1 and X_2. We have that $Min_{\preceq}(X_1 \cup X_2) = A \neq Min_{\preceq}(X_1) \cup Min_{\preceq}(X_2) = A \cup E$. The generalization of conservative upgrade that incorporates a constraint analogous to Cubitt and Sugden's privacy of tie-breaking property should result in the following plausibility ordering:

$$A \cup E \prec B \prec C \cup D \prec F \cup G.$$

The formal definition is:

Definition 6.2.6 (Generalized Conservative Upgrade). Let $\mathcal{M} = \langle W, \preceq, \sigma \rangle$ be a plausibility model and $\{E_1, \ldots, E_k\}$ a set of events. Define $\mathcal{M}^{\uparrow\{E_1,\ldots,E_k\}} = \langle W^{\uparrow\{E_1,\ldots,E_k\}}, \preceq^{\uparrow\{E_1,\ldots,E_k\}}, \sigma^{\uparrow\{E_1,\ldots,E_k\}} \rangle$ as follows: $W^{\uparrow\{E_1,\ldots,E_k\}} = W$, $\sigma^{\uparrow\{E_1,\ldots,E_k\}} = \sigma$ and, if $B = Min_{\preceq}(E_1) \cup Min_{\preceq}(E_2) \cup \cdots \cup Min_{\preceq}(E_k)$, then

1. if $v \in B$, then $v \preceq^{\uparrow\{E_1,\ldots,E_k\}} x$ for all $x \in W$; and
2. for all $x, y \in W - B$, $x \preceq^{\uparrow\{E_1,\ldots,E_k\}} y$ iff $x \preceq y$.

Remark 6.2.7 (Suspending Judgement). A generalized conservative upgrade with $\{E, \overline{E}\}$, where $\overline{E}$ is the complement of E, can be interpreted as a *suspension of judgement* regarding E (cf. Holliday (2009) for a discussion). We do not offer an extended discussion of belief suspension here, but we suggest that a natural response

[9] Rabinovich takes this even further and argues that from the principle of indifference, players must assign equal probability to all choice-worthy options (Rabinowicz 1992).

is to learning that there are more than one chose-worthy action for the players is to *suspend judgement* about which options the relevant players will *pick*.

A generalized conservative upgrade of $\{E_1, \ldots, E_k\}$ "flattens" out the players' beliefs relative to this set of events. After the upgrade, the player will consider each of the E_i equally plausible. But this means that, if w is a most plausible E_i-world and v is a most plausible E_j-world, the player forgets whatever reason she had for considering state w more plausible than v (or vice versa). This suggests a generalization of *radical upgrade* , where the players remember their earlier reasons for considering some states more plausible than others. The idea is to update with a set of events as in Definition 6.2.6, but to maintain the original ordering within the union of the most plausible E_i-worlds.

Definition 6.2.8 (Generalized Radical Upgrade). Let $\mathcal{M} = \langle W, \preceq, \sigma\rangle$ be a plausibility model and $\{E_1, \ldots, E_k\}$ a set of events. Define $\mathcal{M}^{\Uparrow\{E_1,\ldots,E_k\}} = \langle W^{\Uparrow\{E_1,\ldots,E_k\}}, \preceq^{\Uparrow\{E_1,\ldots,E_k\}}, \sigma^{\Uparrow\{E_1,\ldots,E_k\}}\rangle$ as follows: $W^{\Uparrow\{E_1,\ldots,E_k\}} = W$, $\sigma^{\Uparrow\{E_1,\ldots,E_k\}} = \sigma$ and, if $B = Min_{\preceq}(E_1) \cup Min_{\preceq}(E_2) \cup \cdots \cup Min_{\preceq}(E_k)$, then

1. for all $v \in B$, $v \preceq^{\Uparrow\{E_1,\ldots,E_k\}} x$ for all $x \in W - B$;
2. for all $x, y \in B$, $x \preceq^{\Uparrow\{E_1,\ldots,E_k\}} y$ iff $x \preceq y$; and
3. for all $x, y \in W - B$, $x \preceq^{\Uparrow\{E_1,\ldots,E_k\}} y$ iff $x \preceq y$.

Applying this definition to the running example in this section results in the plausbility ordering:

$$A \prec E \prec B \prec C \cup D \prec F \cup G.$$

We will see other examples of the transformations defined above in the next section. These transformations can be logically analyzed using standard techniques from dynamic epistemic/doxastic logic literature (e.g., the "reduction axiom method").

6.3 Rational Deliberation via Iterated Belief Updates

In this section, we use the ideas developed in Sect. 6.2 to formally define our model of deliberation in games. The idea is that a player's "rational response" to a given categorization is to transform the current informational context using one of the transformations from the Sect. 6.2.2. To make this precise, we need to *describe* a categorization.

Definition 6.3.1 (Language for a Game). Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a finite strategic game. Without loss of generality, assume that each of the S_i are disjoint, and let $\mathsf{At}_G = \{P^i_a \mid a \in S_i,\ i \in N\}$ 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 At_G and closed under the boolean connectives $\neg$ and $\wedge$. The other boolean connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) are defined as usual.

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

Definition 6.3.2 (Interpretation of $\mathcal{L}_G$). 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 $[\![\cdot]\!]_{\mathcal{M}_G} : \mathcal{L}_G \rightarrow \wp(W)$ by induction on the structure of $\mathcal{L}_G$ as follows: $[\![P^i_a]\!]_{\mathcal{M}_G} = \{w \mid \sigma_i(w) = a\}$, $[\![\neg\varphi]\!]_{\mathcal{M}_G} = W - [\![\varphi]\!]_{\mathcal{M}_G}$ and $[\![\varphi \wedge \psi]\!]_{\mathcal{M}_G} = [\![\varphi]\!]_{\mathcal{M}_G} \cap [\![\psi]\!]_{\mathcal{M}_G}$.

Let X and Y be two sets of propositions; we define $X \wedge Y := \{\varphi \wedge \psi \mid \varphi \in X, \psi \in Y\}$

Definition 6.3.3 (Describing a categorization). Let G be a finite game and $\mathcal{M}_G$ an informational context of G. Given a categorization $\mathbf{S}(\mathcal{M}_G)$, let $Do(\mathbf{S}(\mathcal{M}_G))$ denote the set of formulas that *describe* $\mathbf{S}$. This set is defined as follows. For each $i \in N$, let:

$$Do_i(\mathbf{S}_i(\mathcal{M}_G)) = \{P^i_a \mid a \in \mathbf{S}^+_i(\mathcal{M}_G)\} \cup \{\neg P^i_b \mid b \in \mathbf{S}^-_i(\mathcal{M}_G)\}.$$

Then, define $Do(\mathbf{S}(\mathcal{M}_G)) = Do_i(\mathbf{S}_i(\mathcal{M}_G)) \bigwedge Do_2(\mathbf{S}_2(\mathcal{M}_G)) \cdots \bigwedge Do_n(\mathbf{S}_n(\mathcal{M}_G))$.

The general project is to understand the interaction between types of categorizations (e.g., choice rules) and types of model transformations (representing the rational deliberation process). One key question, is whether (and under what conditions) a deliberation process *stabilizes*? There are a number of ways to make precise what it means to stabilize (see Baltag and Smets (2009) for a discussion).

Definition 6.3.4 (Stable in Beliefs). Suppose that $\mathcal{M} = \langle W, \preceq, \sigma \rangle$ and $\mathcal{M}' = \langle W, \preceq', \sigma' \rangle$ are two plausibility models based on the same set of states.[10] We say that $\mathcal{M}$ and $\mathcal{M}'$ are **stable with respect to the players' beliefs** if the set of propositions that are believed in $\mathcal{M}$ is the same as those believed in $\mathcal{M}'$. Equivalently, $\mathcal{M}$ and $\mathcal{M}'$ are stable with respect to beliefs provided $Min_{\preceq}(W) = Min_{\preceq'}(W)$. We write $\mathcal{M} \equiv_B \mathcal{M}'$ when $\mathcal{M}$ and $\mathcal{M}'$ are stable with respect to beliefs.

In this paper, it is enough to define stabilization in terms of the players' simple beliefs because, during the deliberation process, we incorporate only information about what the players are going to do (as opposed to higher-order information[11]). We are now ready to formally define a "deliberation sequence":

[10] So, we assume that the models agree about which outcomes of the game have not been ruled out.

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

Definition 6.3.5 (Upgrade Sequence). Given a game G and an informational context $\mathcal{M}_G$, an upgrade sequence of type τ, induced by $\mathcal{M}_G$ is a 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 $\mathcal{M}_n \equiv_B \mathcal{M}_{n+1}$. The next section has a number of examples of upgrade sequences, some that stabilize and others that do not stabilize.

In the remainder of this section, we discuss a number of abstract principles that any upgrade sequence should satisfy. To state these properties, we need some notation. Let U be a fixed set of states and G a fixed strategic game. We restrict attention to transformations between models of G whose states come from the same set of states U. Let $\mathbb{M}_G$ be the set of all such plausibility models. A model transformation is 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 functions τ make sense, given that we intend τ to model belief changes as the players deliberate about what to do. The first set of principles ensure that the categorizations are "sensitive" to the players' beliefs and that the players respond to the categorizations in the appropriate way. Suppose that $\mathcal{X} = \{\varphi_1, \ldots, \varphi_k\}$ is a finite set of $\mathcal{L}_G$ formulas and $\mathcal{M} \in \mathbb{M}_G$.

A1 The operation τ depends only on the truth set of the formulas: If, for each $i = 1, \ldots, k$, $[\![\varphi_i]\!]_\mathcal{M} = [\![\psi_i]\!]_\mathcal{M}$, then $\tau(\mathcal{M}, \mathcal{X}) = \tau(\mathcal{M}, \{\psi_1, \ldots, \psi_n\})$.

A2 The operation τ is idempotent[12] in the language $\mathcal{L}_G$: $\tau(\mathcal{M}, \mathcal{X}) = \tau(\mathcal{M}^{\tau(\mathcal{X})}, \mathcal{X})$.

Property **A1** says that the belief transformations depend only on the propositions expressed by a formula by treating equivalent formulas the same way. The second property **A2** says that receiving the exact same information twice does not have any effect on the players' beliefs. These are natural properties that are satisfied by any belief-change policy. Certainly, there may be other properties that one may want to impose (for example, variants of the AGM postulates Alchourrón et al. 1985). We leave a discussion of additional principles for another paper. The next two properties ensure that the transformation responds "properly" to a categorization.

A3 For all models $\mathcal{M}, \mathcal{M}' \in \mathbb{M}_G$ and categorizations $\mathbf{S}$, if $\mathcal{M} \equiv_B \mathcal{M}'$, then $\mathbf{S}(\mathcal{M}) = \mathbf{S}(\mathcal{M}')$.

A4 For all models $\mathcal{M}, \mathcal{M}' \in \mathbb{M}_G$, $\tau(\mathcal{M}, Do(\mathbf{S}(\mathcal{M})))$ incorporates $\mathbf{S}(\mathcal{M})$.

[12] Here, it is crucial that the language $\mathcal{L}_G$ does not contain any modalities.

Property **A3** guarantees that the categorization depends only on the players' beliefs. Property **A4** ensures the players are responding to the categorizations in the right way. The properties **A1**, **A2**, **A3** and **A4** are the minimal set of principles that an upgrade sequence must satisfy in order to serve as a model of deliberation in games. We conclude this section by discussing conditions that guarantee that an upgrade sequence will stabilize.

There are two main reasons why an upgrade sequence would stabilize. The first is due to the properties of the transformation (for example, it is clear that upgrade streams with public announcements always stabilize). The second is because the choice rule satisfies a monotonicity property so that, eventually, the categorizations stabilize, and so, there is no new information to change the plausibility ordering. One way to guarantee that upgrade sequences stabilize is to assume that the categorizations satisfy a monotonicity property.

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 $i \in N$, $\mathbf{S}_i^+(\mathcal{M}_n) \subseteq \mathbf{S}_i^+(\mathcal{M}_{n+1})$.

Property **Mon**$^-$ means that once an option for a player is classified as "not rationally permissible", it cannot, at a later stage of the deliberation process, drop this classification. Property **Mon**$^+$ says that either the rationally permissible options satisfy the same monotonicity property or they are completely determined by the set of rationally impermissible options.

Theorem 6.3.6. *Suppose that G is a finite strategic game and that all of the above properties are satisfied. Then, every upgrade sequence $(\mathcal{M}_n)_{n\in\mathbb{N}}$ for G stabilizes.*

Proof. Let $G = \langle N, \{S_i\}_{i\in N}, \{u_i\}_{i\in N}\rangle$ be a finite strategic game. By properties **Mon**$^-$ and **Mon**$^+$ we have either for all upgrade streams $(\mathcal{M}_n)_{n\in\mathbb{N}}$ and players $i \in N$,

1. $\mathbf{S}_i^-(\mathcal{M}_0) \subseteq \mathbf{S}_i^-(\mathcal{M}_1) \subseteq \cdots \mathbf{S}_i^-(\mathcal{M}_n) \subseteq \cdots$ is an infinitely increasing sequence of subsets of S_i and $\mathbf{S}_i^+(\mathcal{M}_0) \supseteq \mathbf{S}_i^+(\mathcal{M}_1) \supseteq \cdots \mathbf{S}_i^+(\mathcal{M}_n) \supseteq \cdots$ is an infinite decreasing sequence of subsets of S_i; or
2. Both,

$$\mathbf{S}_i^-(\mathcal{M}_0) \subseteq \mathbf{S}_i^-(\mathcal{M}_1) \subseteq \cdots \mathbf{S}_i^-(\mathcal{M}_n) \subseteq \cdots$$

and

$$\mathbf{S}_i^+(\mathcal{M}_0) \subseteq \mathbf{S}_i^+(\mathcal{M}_1) \subseteq \cdots \mathbf{S}_i^+(\mathcal{M}_n) \subseteq \cdots$$

are infinite increasing sequences of subsets of S_i.

Since each S_i is assumed to be finite, for each player i, there is an n_i such that $\mathbf{S}_i^-(\mathcal{M}_{n_i}) = \mathbf{S}_i^-(\mathcal{M}_{n_i+i})$ and $\mathbf{S}_i^+(\mathcal{M}_{n_i}) = \mathbf{S}_i^+(\mathcal{M}_{n_i+i})$. Let m be the maximum of $\{n_i \mid i \in N\}$. Then, we have $\mathbf{S}(\mathcal{M}_m) = \mathbf{S}(\mathcal{M}_{m+1})$. All that remains is to show that for all $x > m$, $\mathcal{M}_x = \tau(\mathcal{M}_x)$. This follows by an easy induction on x. The key calculation is: for each $x \in \mathbb{N}$, let $\mathcal{D}_x$ be the appropriate description of $\mathbf{S}(\mathcal{M}_x)$.

$$\begin{aligned}\mathcal{M}_{m+2} = \tau(\mathcal{M}_{m+1}, \mathcal{D}_{m+1}) &= \tau(\mathcal{M}_m^{\tau(\mathcal{D}_m)}, \mathcal{D}_{m+1}) \\ &= \tau(\mathcal{M}_m^{\tau(\mathcal{D}_m)}, \mathcal{D}_m) \text{ (since } \mathbf{S}(\mathcal{M}_m) = \mathbf{S}(\mathcal{M}_{m+1})) \\ &= \tau(\mathcal{M}_m, \mathcal{D}_m) = \mathcal{M}_{m+1}\end{aligned}$$

This concludes the proof. QED

A number of researchers have noticed that monotonicity of the choice rule is important for an epistemic analysis of games (see Apt and Zvesper (2010b) for a discussion). An immediate corollary of Theorem 6.3.6 is:

Corollary 6.3.7. *If the categorization method is strict dominance, then any upgrade sequence of type τ stabilizes, where τ is any of the transformations discussed in this paper (e.g., public announcement, (generalized) radical upgrade and (generalized) conservative upgrade).*

This is related to van Benthem's iterated "soft" announcements of rationality (van Benthem 2007) and Apt and Zvesper's results about stabilization of beliefs in games (Apt and Zvesper 2010a).

6.4 Case Study: Iterated Weak Dominance

Larry Samuelson (1992) points out an explicit puzzle surrounding the epistemic foundations of iterated removal of weakly dominated strategies (IEWDS) – also known as the IA solution. He shows (among other things) that there is no epistemic model of the following game with at least one state satisfying "common knowledge of admissibility" (i.e., a state in which there is common knowledge that the players do not play a strategy that is weakly dominated).

		Bob	
		L	R
Ann	u	1, 1	1, 0
	d	1, 0	0, 1

In the above game, d is weakly dominated by u for Ann. If Bob believes that Ann is rational (in the sense that she will not choose a weakly dominated strategy), then he can conclude that u is more plausible than d. In the smaller game, action R is now strictly dominated by L for Bob. If Ann believes that Bob is rational and that Bob knows that she is rational (and thus, d is rationally impermissible), then she can conclude that L is more plausible than R. Assuming that the above reasoning is transparent to both Ann and Bob, it is common knowledge that Ann will play u and Bob will play L. But now, what is the reason for Bob to rule out the possibility that

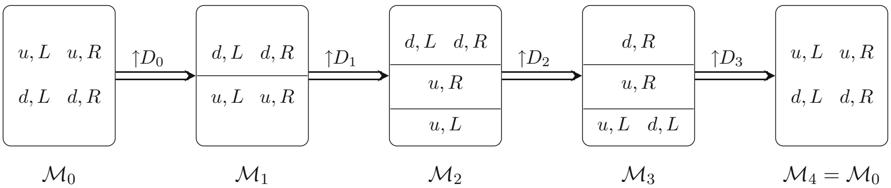
Ann will play d? He believes that Ann believes that he is going to play L, and both u and d are best responses to L.

The general framework introduced in Sects. 6.2 and 6.3 offers a new, dynamic perspective on Samuelson's analysis, as well as on reasoning with weak dominance more generally. Note that we are not providing an *alternative* epistemic characterization of IEWDS. Both Brandenburger et al. (2008) and Halpern and Pass (2009) have convincing results here. Our goal is to use this solution concept to illustrate our general approach.

Generalized Conservative Upgrade with Weak Dominance Dynamically, Samuelson's analysis of the above game corresponds to non-stabilization of an upgrade sequence. The players are not able to reason their way to stable, common belief in admissibility. To capture this intuition, in light of Theorem 6.3.6, we need to work with a non-monotonic categorization. Before stating the observation formally, we need one more definition. A **full model** of a game G is one in which all outcomes of the game are in the model (i.e., for any profile $\vec{s}$, there is a state w satisfying $\sigma(w) = \vec{s}$) and the states are all equally plausible.

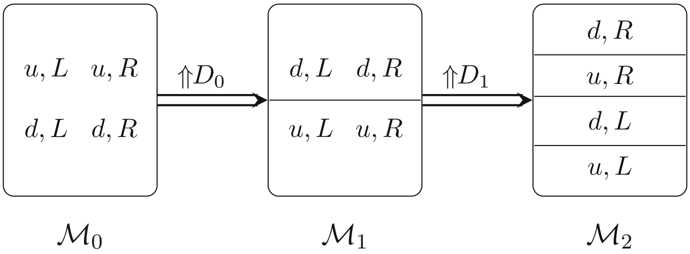
Observation 6.4.7 *Starting with the initial full model of the above game, the conservative upgrade sequence for conservative upgrade and weak dominance does not stabilize.*

The proof of this Observation is provided by the following looping stream of conservative upgrades:



where for $i = 1, 2, 3, 4$, $D_i = Do(\mathbf{WD}(\mathcal{M}_i))$. Intuitively, from $\mathcal{M}_0$ to $\mathcal{M}_2$ the players have reasons to exclude d and R, leading them to commonly believe that u, L is played. At that stage, however, d is admissible for Ann, canceling the players' reason for ruling out this strategy. The rational response 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 $\mathcal{M}_0$. This process loops forever; the agents' reasoning does not stabilize.

Generalize Radical Upgrade with Weak Dominance Generalized radical upgrade stabilizes plain beliefs even for non-monotonic choice rules such as weak dominance. Consider, again, Samuelson's game given above. Starting with the full model of this game, the upgrade stream stabilizes on a model with the (common) belief that all the players will play the IEWDS outcome.



where D_0 and D_1 are as above. Intuitively, what happens is the following: Just as with conservative upgrade, $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively, give the agents reasons to believe that Ann will not play d, and that Bob will not play R. This leads to $\mathcal{M}_2$, where, like before, d is admissible given that Ann believes that Bob will play L. Radical upgrade, however, does not allow this fact to override her reason for not playing d: her rational response is to rank u, L and d, L above all other possible outcomes, but to keep the relative ordering of these two, reflecting the fact that she previously ruled out u.

Stabilization of radical upgrade puts Samuelson's observation into perspective. Such an upgrade forces the agents to remember the reasons they had earlier in the deliberation. Previous reasons constrain the domain of permissibility at later stages in the deliberation process. What is permissible for Ann at $\mathcal{M}_2$ depends on the deliberation process that led to this model, and, in particular, on the existence of an (earlier) reason not to play d. This was not the case for conservative upgrade. Reasons at each stage were evaluated de novo, without reference to the reasoning history. This is what led the upgrade sequence for Samuelson's game into looping, to the "paradox" of admissibility. We leave open for discussion whether this constitutes an argument to the effect that players "should" keep track of their reasons while reasoning to a specific informational context. For now, we content ourselves with the observation that there is a tight connection, on the one hand, between remembering one's reasons and stabilization of reasoning under admissibility and, on the other hand, between letting new reasons override previous ones and the possibility of never-ending reasoning chains.

6.5 Concluding Remarks

A general theory of rational deliberation for game and decision theory is a big topic, and, thus, it is beyond the scope of this article to discuss the many different aspects and competing perspectives on such a theory. The reader is referred to Brian Skyrms' (1990, Chap.7) for a broader discussion. The main contribution of this paper is to lay the foundation for a formal theory of deliberation in games, based on recent work on dynamic logics of knowledge and belief. We focused on one specific question: What type of process can be used to *generate* a game model?

The most pressing philosophical issue concerns the role that a theory of deliberation plays in rational choice theory (cf. Levi 1993; Rabinowicz 2002; Schick 1979; Arntzenius 2008). On the technical side, throughout the paper, we worked with (logical) models of *all out* attitudes, leaving aside probabilistic, graded beliefs, even though they are arguably the most widely used in the current literature on epistemic foundations of game theory. It is an important, and non-trivial, task to transpose the dynamic perspective on informational contexts that we advocate here to such probabilistic models. We leave that for future work.

Finally, we should stress that the dynamic perspective on informational contexts is a natural complement, and not an alternative, to existing epistemic characterizations of solution concepts (van Benthem et al. 2011). Epistemic characterizations of solution concepts offer rich insights into the consequences of taking the informational contexts of strategic interaction seriously. What we proposed here is a first step towards understanding how and why such a context might arise.

References

Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: partial meet contraction and revision functions. J. Symb. Log. **50**, 510–530 (1985)

Anglberger, A.J.J., Gratzl, N., Roy, O.: Obligation, free choice and the logic of weakest permission. Rev. Symb. Log. **8**(4), 807–827 (2015)

Apt, K., Zvesper, J.: Public announcements in strategic games with arbitrary strategy sets. In: Proceedings of LOFT 2010, Toulouse, France (2010a)

Apt, K., Zvesper, J.: The role of monotonicity in the epistemic analysis of strategic games. Games **1**(4), 381–394 (2010b)

Arntzenius, F.: No regrest, or: edith piaf revamps decision theory. Erkenntnis **68**, 277–297 (2008)

Baltag, A., Smets, S.: Group belief dynamics under iterated revision: fixed points and cycles of joint upgrades. In: Proceedings of Theoretical Aspects of Rationality and Knowledge, Stanford (2009)

Baltag, A., Smets, S., Zvesper, J.: Keep 'hoping' for rationality: a solution to the backwards induction paradox. Synthese **169**, 301–333 (2009)

Board, O.: Dynamic interactive epistemology. Games Econ. Behav. **49**, 49–80 (2004)

Boutilier, C.: Conditional logics for default reasoning and belief revision. PhD thesis, University of Toronto (1992)

Brandenburger, A., Friedenberg, A., Keisler, H.J.: Admissibility in games. Econometrica **76**, 307–352 (2008)

Cubitt, R., Sugden, R.: Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. Econ. Philos. **19**(2), 175–210 (2003)

Cubitt, R., Sugden, R.: The reasoning-based expected utility procedure. Games Econ. Behav. **71**(2), 328–338 (2011)

Cubitt, R., Sugden, R.: Common reasoning in games: a Lewisian analysis of common knowledge of rationality. Econ. Philos. **30**(3), 285–329 (2014)

Dekel, E., Siniscalchi, M.: Epistemic game theory. In: Peyton Young, H., Shmuel, Z. (eds.) Handbook of Game Theory with Economic Applications, vol. 4, pp. 619–702. Elsevier, Amsterdam (2015)

Gerbrandy, J.: Bisimulations on planet Kripke. PhD thesis, University of Amsterdam (1999)

Halpern, J., Pass, R.: A logical characterization of iterated admissibility. In: Heifetz, A. (ed.) Proceedings of the Twelfth Conference on Theoretical Aspects of Rationality and Knoweldge, Stanford, pp. 146–155 (2009)

Harsanyi, J.: The tracing procedure: a Bayesian approach to defining a solution for *n*-person noncooperative games. Int. J. Game Theory **4**, 61–94 (1975)

Holliday, W.: Trust and the dynamics of testimony. In: Logic and interaction rationality: Seminar's yearbook 2009, pp. 147–178. ILLC technical reports (2009)

Lamarre, P., Shoham, Y.: Knowledge, certainty, belief and conditionalisation. In: Proceedings of the International Conference on Knoweldge Representation and Reasoning, pp. 415–424 (1994)

Levi, I.: Rationality, prediction, and autonomous choice. Can. J. Philos. **23**(1), 339–363 (1993)

Lewis, D.: Convention. Harvard University Press, Cambridge (1969)

Lewis, D.: Counterfactuals. Blackwell, Oxford (1973)

Leyton-Brown, K., Shoham, Y.: Essentials of Game Theory: A Concise Multidisciplinary Introduction. Morgan & Claypool, San Rafael (2008)

Pacuit, E.: Dynamic epistemic logic II: logics of informaiton change. Philos. Compass **8**(9), 815–833 (2013)

Pacuit, E.: Dynamic models of rational deliberation in games. In: van Benthem, J., Ghosh, S., Verbrugge, R. (eds.) Models of Strategic Reasoning: Logics, Games and Communities. Springer (2015)

Pacuit, E., Roy, O.: Epistemic foundations of game theory. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (2015)

Perea, A.: Epistemic Game Theory: Reasoning and Choice. Cambridge University Press, New York (2012)

Peterson, M.: An Introduction to Decision Theory. Cambridge University Press, Cambridge/New York (2009)

Plaza, J.: Logics of public communications. In: Emrich, M.L., Pfeifer, M.S., Hadzikadic, M., Ras, Z.W. (eds.) Proceedings, 4th International Symposium on Methodologies for Intelligent Systems, pp. 201–216 (1989)

Rabinowicz, W.: Tortous labyrinth: noncooperative normal-form games between hyperrational players. In: Bicchieri, C., Chiara, M.L.D. (eds.) Knowledge, Belief and Strategic Interaction. Cambridge University Press, Cambridge/New York, pp. 107–125 (1992)

Rabinowicz, W.: Does practical deliberation crowd out self-prediction. Erkenntnis **57**, 91–122 (2002)

Rott, H.: Shifting priorities: simple representations for 27 itereated theory change operators. In: Lagerlund, H., Lindström, S., Sliwinski, R. (eds.) Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg. Uppsala Philosophical Studies, vol. 53, pp. 359–384 (2006)

Roy, O., Anglberger, A.J.J., Gratzl, N.: The logic of best action from a deontic perspective. In: Baltag, A., Smets, S. (eds.) Johan FAK van Benthem on Logical and Informational Dynamics. Springer (2014)

Samuelson, L.: Dominated strategies and common knowledge. Game Econ. Behav. **4**, 284–313 (1992)

Schick, F.: Self-knowledge, uncertainty and choice. Br. J. Philos. Sci. **30**(3), 235–252 (1979)

Skyrms, B.: The Dynamics of Rational Deliberation. Harvard University Press, Cambridge (1990)

Ullmann-Margalit, E., Morgenbesser, S.: Picking and choosing. Soc. Res. **44**, 757–785 (1977)

van Benthem, J.: Dynamic logic for belief revision. J. Appl. Non-class. Log. **14**(2), 129–155 (2004)

van Benthem, J.: Rational dynamics and epistemic logic in games. Int. Game Theory Rev. **9**(1), 13–45 (2007)

van Benthem, J.: Logical Dynamics of Information and Interaction. Cambridge University Press (2010)

van Benthem, J.: Logic in Games. MIT, Cambridge/London (2014)

van Benthem, J., Gheerbrant, A.: Game solution, epistemic dynamics and fixed-point logics. Fund. Inf. **100**, 1–23 (2010)

van Benthem, J., Pacuit, E., Roy, O.: Towards a theory of play: a logical perspective on games and interaction. Games **2**(1), 52–86 (2011)