An abstract approach to reasoning about games with mistaken and changing beliefs

Benedikt Löwe[1,2,3], Eric Pacuit[1⋆]

[1] Institute for Logic, Language and Computation, Universiteit van Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
{bloewe,epacuit}@science.uva.nl

[2] Department Mathematik, Universität Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

[3] Mathematisches Institut, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 1, 53115 Bonn, Germany

1 Introduction

Both logic and game theory have developed an interest in analyzing what constitutes rational behaviour under uncertainty. One particular interesting encounter between logic and game theory is the use of belief revision techniques in the sense of [Gä92] as a means of analysis of games.[1]

The game-theoretic analysis of rationality and the study of belief revision have in common that they have a normative hue; they are (overly) concerned with questions of what constitutes rational behaviour and what would be quality measures for rationality. On the purely logical side, without taking into account context, this is a doomed enterprise: if you believe p and $p \to q$, and learn for a fact that $\neg q$ holds, then whether you give up p or the implication will depend on what these statements are and what the context is.[2]

In this paper, we present an approach to belief revision in games that renounces any claims of normativity. We shall present an account of analysis of games in which all agents have beliefs about the preferences of the other agents and beliefs about those beliefs, and so on. Based on what happens in the game, they can change their beliefs, and again, these changes can be the subject of beliefs of the agents. Our account is purely formal and does not presuppose any theory of rationality or of belief change. As a consequence, our account is able to properly describe also bizarre and irrational behaviour.[3]

⋆ The authors would like to thank Andrès Perea (Maastricht): The syntax described in this note is based on the one developed in [LöPa$_0$Pe∞].

[1] The following is a list of relevant papers combining epistemic logic and game theory: [Au99], [St98], [Bo$_1$Ba99], [Br06], [Mo+97], [dB04]. More specifically, the papers [St98], [Bo$_0$04], and [Pe05] have pointed out the importance of belief revision in the context of reasoning about solution concepts in games.

[2] *Cf.* [Lö06] for a related discussion.

[3] It is a well-known feature of artificial intelligence that what is bizarre if you think of human agents can be perfectly fine for artificial agents. As an example, think of

In § 2, we give an example story of mistaken beliefs that describes how we imagine our formal setting to be applied. Then, in § 3, we give the formal definitions of our syntax and semantics and how to use this setting to get a backwards induction solution. In the following § 4, we then apply our semantics to the story related in § 2 and give an analysis of it. Finally, in § 5 we list projects for the future based on the framework presented in this paper.

2 A story of reasoning with false beliefs

The following is a fictitious story in the style of a TV drama. The reader can imagine that this is the outline of a script. The reasoning processes referred to in the story can be made visible to the audience by monologues (Walter talking to himself in his car) or by conversations with some *confidant* or *confidante*.

> *Sue and Jeff have known each other for years. They studied computer science together in the 1980s, and both started their own software companies in the 1990s. Sue is married to Walter, an artist, and Jeff is married to Mary. In the past years, Sue and Mary have become best friends.*
> *However, unbeknownst to Sue, her husband Walter and Mary have an affair. Walter, being absolutely dependant on the money of his wife, has no intention of leaving her at all, and wants to avoid that she gets to know about this affair at all costs. He believes that the fact that Sue and Mary are best friends acts as a safeguard for his secret: Mary must be aware that she will lose Sue as a friend if Sue finds out, and Mary clearly doesn't want that. So, Walter convinced himself that Mary will never ask him to separate from Sue or –even worse– tell Sue about the affair.*
> *Mary on the other hand is rather unhappy with Jeff, and really wants to leave him. She believes that her friendship with Sue is robust enough to survive the fact that she has an affair with Walter. In her dreams, she imagines a nice future with Walter. She is convinced that if she presses Walter enough, he will finally leave Sue for her. She can make up with Sue afterwards.*
> *One morning, she gives Walter an ultimatum: he should make up his mind and choose between her and Sue. Walter is ultimately confused: he must have misjudged Mary. Stuck in the traffic jam on his way to an appointment with a potential client, his mind raced: If he chooses Mary, then Sue would know about their affair, and Mary would lose her best friend. What was Mary thinking? The only rational explanation that he could come up with was that Mary wanted to be with him so badly that she would give up her friendship with Sue for it... Obviously, Walter couldn't leave Sue for financial reasons. But he needed to be careful here:*

computational social choice: for human beings, we would hardly call "maximize the benefit for the strongest agent" a reasonable social procedure; in artificial situations, this may very well be desirable.

if he said no to Mary, would she tell Sue? No, he reasoned, since then she would lose both Sue and him which is definitely worse than just losing him. So, he'd be safe. Smiling, he used his cell phone to call Mary and tell her that he would not leave Sue.
When she hung up the phone, Mary was fuming with anger. Apparently, Walter wanted to stay with Sue. "Well, if that's what he wants, then I am not interested in him anymore. I should cut my losses, and at least be honest to my best friend," she reasoned, and acted accordingly.
And Mary was right in her judgement of Sue. The two women discussed the matter, and when Walter returned from his appointment in the afternoon, his paintings were standing on the front lawn of their house and the lock of the front door had been changed. Walter gazed emptily at his paintings searching for a logician to help him to figure out what had happened.

We should stress that human beings have no problems in analysing an episode like this – with ease, they can make judgements like "Walter is wrong about his judgement of Sue and Mary; after the ultimatum, there was no chance of staying together with Sue anymore, but he could have saved his relationship with Mary hadn't he misjudged his wife".

For computational situations, we would like to be able to do the same within some formal system. Being able to formally access the reasoning structure of episodes like this is crucial for the analysis of games with mistaken beliefs.

3 Formalization

In the following, we shall describe a formalization of reasoning about mistaken beliefs and belief change that goes back to a more elaborate version presented in [LöPa$_0$Pe∞].

Let I be the finite set of players. A tree T is a finite set of nodes together with an edge relation (in which any two nodes are connected by exactly one path). Let root_T denote the root of the tree and $\mathbf{tn}(T)$ denote the set of terminal nodes of T. We write $t \in T$ if t is a node in the tree T. If $t \in T$, let $\mathbf{succ}_T(t)$ denote the set of immediate T-successors of T.

An **extensive game form** is a tuple $\langle I, T, \mu \rangle$ where I is a set of players, T is a tree and μ is a **moving function**. That is,

$$\mu : T \backslash \mathbf{tn}(T) \to I,$$

where, intuitively, if $\mu(T) = i$ then it is i's move at node t. We call total orders $\succeq$ on $\mathbf{tn}(T)$ **preferences** or **preference relations**. If $\langle I, T, \mu \rangle$ is an extensive game form, and $\{\succeq_i ; i \in I\}$ is an assignment of preference relations to the players (if $t_1 \succeq_i t_2$, we say "player i prefers the node t_1 over node t_2"), then we call

$$\langle I, T, \mu, \{\succeq_i ; i \in I\} \rangle$$

an **extensive game**. This model of a game is completely standard and discussions can be found in any game theory text (for example, *cf.* [OsRu94]).

From now on, we fix an extensive game form $\mathfrak{G} = \langle I, T, \mu \rangle$. Let $\mathcal{P}$ be a countable set of **preference symbols**. We typically denote elements of $\mathcal{P}$ by $\sqsupseteq$ and interpret them with preferences. For $i \in I$ and $\sqsupseteq \in \mathcal{P}$, the string $[i, \sqsupseteq]$ is called an **atomic formula** whose intended interpretation is "player i has the preference denoted by $\sqsupseteq$". The set of atomic formulae is denoted by At. A set $D \subseteq \mathsf{At}$ is called a **description** if for each $i \in I$, there is a unique preference symbol $\sqsupseteq$ such that $[i, \sqsupseteq] \in D$. As we want to add belief and belief change to our language, we add modal operators $\Box_i^t$ for every $t \in T$ and $i \in I$. The intended meaning $\Box_i^t \varphi$ is "if the game reached node t, then player i believes φ".

The **state language** $\mathcal{L}$ is now the closure of At under the operators $\Box_i^t$. A set of formulae S is called a **state** if for each finite sequence of operators $\langle O_0, \ldots, O_u \rangle$ (including the empty sequence) there is a unique description D such that for each $\varphi \in \mathsf{At}$,

$$\varphi \in D \text{ if and only if } O_0 \cdots O_u \varphi \in S.$$

We denote the set of all states with $\mathbb{S}$. Given a state S, an $i \in I$ and a $t \in T$, let $S_i^t = \{\varphi \ ; \ \Box_i^t \varphi \in S\}$ ("agents i's beliefs at t in state S"). Obviously, if S is a state, then S_i^t is a state.

An **interpretation** $\mathfrak{I}$ is any function assigning preferences to elements of $\mathcal{P}$.[4] A **game model** based on $\mathfrak{G}$ is a tuple $\langle \mathfrak{G}, S, \mathfrak{I} \rangle$ where $\mathfrak{I}$ is an interpretation and $S \in \mathbb{S}$ is the initial state. The initial state and the interpretation contain information about the true preferences of players at the beginning of the game: player i has preference $\mathfrak{I}(\sqsupseteq)$ if and only if $[i, \sqsupseteq] \in S$.

Given a game model $\langle \mathfrak{G}, S, \mathfrak{I} \rangle$, we can now fully analyze the game and predict its outcome (assuming that the players follow the backwards induction solution).

We need some preliminary definitions: For a finite tree T, we shall call a function $\ell : I \times T \setminus \{\mathsf{root}_T\} \to \mathbf{tn}(T)$ a **labelling**, and a function $\ell : I \times T \to \mathbf{tn}(T)$ a **full labelling**. We want to associate to each state S a labelling ℓ_S such that intuitively, the following holds: "if S is the true state, then player i believes that if the game reaches node t, then it will end in $\ell_S(i, t)$".

If a tree T has a labelling ℓ_U for every state U and S is a fixed state, then we can define the **S-true run** recursively as follows: let $\succeq_i$ be the true preference of player i, *i.e.*, $[i, \sqsupseteq] \in S$ with $\mathfrak{I}(\sqsupseteq) = \succeq_i$. Then $\mathsf{t}_{T,S}^{[0]} := \mathsf{root}_T$; for a given

[4] Note that our semantics is still very abstract in that it is not taking into account any commonsense properties of belief or rationality. For instance, our definition of state allows that $[i, \sqsupseteq] \in S$, but $\Box_i^{\mathsf{root}_T}[i, \sqsupseteq] \notin S$, *i.e.*, player i has preferences that he doesn't believe he has, or $\Box_i^t \varphi \in S$, but $\Box_i^t \Box_i^t \varphi \notin S$, *i.e.*, a violation of positive introspection, or $\Box_i^t [j, \sqsupseteq]$, for some position t inconsistent with j having the preference $\mathfrak{I}(\sqsupseteq)$, *i.e.*, an irrational belief revision at t. All of these properties conceivably might be useful in some applications, and can easily be excluded by additional axioms if they are not.

$\mathsf{t}^{[k]}_{T,S}$ that is not a terminal node, we let $i := \mu(\mathsf{t}^{[k]}_{T,S})$, and consider $\{\ell_S(i,t)\,;\, t \in \mathbf{succ}(\mathsf{t}^{[k]}_{T,S})\}$. There is a unique $t \in \mathbf{succ}(\mathsf{t}^{[k]}_{T,S})$ associated to the $\succeq_i$-maximal element of that set. Let $\mathsf{t}^{[k+1]}_{T,S} := t$. Since T is finite, there will be some k such that $\mathsf{t}^{[k]}_{T,S}$ is a terminal node. Then $\langle \mathsf{t}^{[0]}_{T,S}, ..., \mathsf{t}^{[k]}_{T,S}\rangle$ is the true run of the game in state S.

Similarly, if a tree T has a labelling ℓ_U for every state U and S is a fixed state, then we can define the **S-subjective run for player j** recursively as follows: let $\succeq^t_{j,i}$ be $\mathfrak{I}(\sqsupseteq)$ for the unique $\sqsupseteq$ such that $[i,\sqsupseteq] \in S^t_j$. Then $\mathsf{t}^{[0]}_{T,j,S} := \mathsf{root}_T$; for a given $\mathsf{t}^{[k]}_{T,j,S}$ that is not a terminal node, we let $i := \mu(\mathsf{t}^{[k]}_{T,j,S})$, and consider $\{\ell_{S^t_j}(i,t)\,;\, t \in \mathbf{succ}(\mathsf{t}^{[k]}_{T,j,S})\}$. As before, let $\mathsf{t}^{[k+1]}_{T,j,S}$ be the unique $t \in \mathbf{succ}(\mathsf{t}^{[k]}_{T,j,S})$ associated to the $\succeq^{\mathsf{t}^{[k]}_{T,j,S}}_{j,i}$-maximal element of the mentioned set. If k is least such that $\mathsf{t}^{[k]}_{T,j,S}$ is a terminal node, then $\langle \mathsf{t}^{[0]}_{T,j,S}, ..., \mathsf{t}^{[k]}_{T,j,S}\rangle$ is the subjective run according to player j of the game in state S.[5]

In our recursive (backwards induction) definitions of the labellings, for every terminal node $t \in \mathbf{tn}(T)$ and arbitrary $S \in \mathbb{S}$ and $i \in I$, we let $\ell_S(i,t) := t$.

To complete the recursion, we take a tree T with $\mathbf{succ}(\mathsf{root}_T) = \{t_0, ..., t_N\}$. We denote by T_n the subtree of T with $\mathsf{root}_{T_n} = t_n$ (for $n \leq N$), and assume that for every state U, there is a full labelling ℓ^n_U on T_n. Fix $S \in \mathbb{S}$ and define a full labelling ℓ_S on T as follows: Clearly, the full labellings ℓ^n_S combine to give a labelling on T; let us call this labelling ℓ_S. We need to define $\ell_S(j, \mathsf{root}_T)$ to make this into a full labelling. For fixed $j \in I$, let $\langle \mathsf{t}^{[0]}_{T,j,S}, ..., \mathsf{t}^{[k]}_{T,j,S}\rangle$ be the subjective run according to player j of the game in state S. Then

$$\ell_S(j, \mathsf{root}_T) := \mathsf{t}^{[k]}_{T,j,S}.$$

Given that S is the true state at the beginning of the game and that ℓ_S is defined recursively as above, we can now analyze the game by means of the S-true run: this will be the actual sequence of events in the game.

4 Analysis of the story

In our story about Walter, Mary and Sue, there are three relevant players $\{\mathbf{W}, \mathbf{M}, \mathbf{S}\}$. The game tree is given in Figure 1. We use the notation ☼ to indicate that a relationship is intact and ⛈ to indicate that it is ended, and list the terminal nodes by the status of the relationships in the order **Sue-Walter**, **Mary-Walter** and **Mary-Sue**; for instance, ⛈☼⛈ stands for "Walter left Sue, is together with Mary, and Sue hates Mary".

[5] Note that the S-subjective run for player j is not necessarily the $S^{\mathsf{root}_T}_j$-true run, as player j's beliefs about the preferences might change from the ones he has in node root_T.

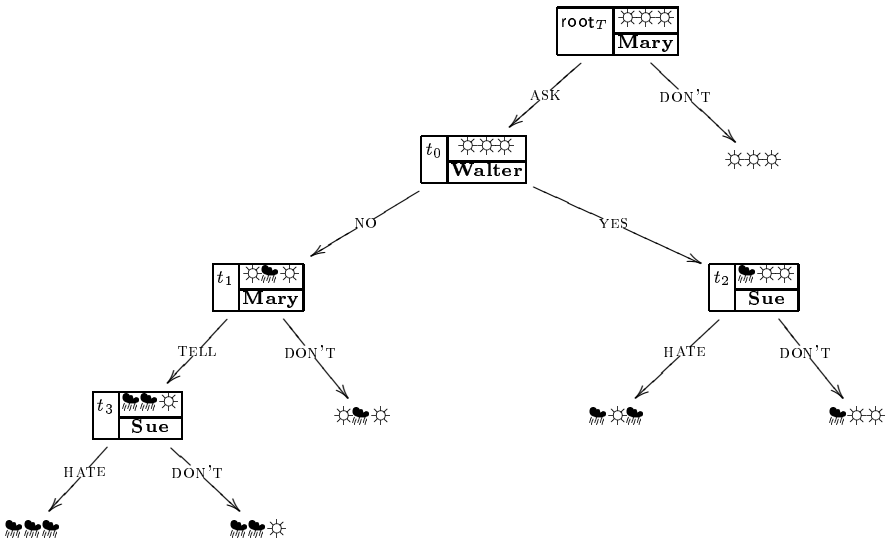
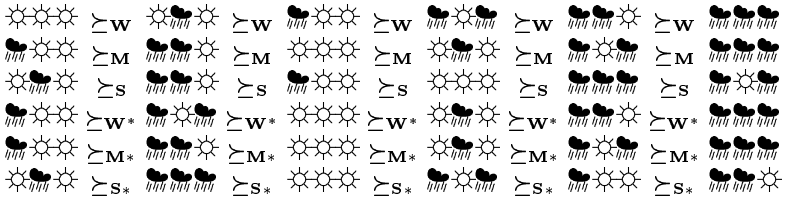


Fig. 1. The game tree for the story about Walter, Mary and Sue.

The only preferences relevant for our game analysis are the following five preferences $\succeq_{\mathbf{W}}$, $\succeq_{\mathbf{M}}$, $\succeq_{\mathbf{S}}$, $\succeq_{\mathbf{W}*}$, $\succeq_{\mathbf{S}*}$, and $\succeq_{\mathbf{M}*}$:



The true preferences of Walter, Mary and Sue are $\succeq_{\mathbf{W}}$, $\succeq_{\mathbf{M}}$, and $\succeq_{\mathbf{S}}$, respectively. The following are the relevant beliefs and belief changes that we put into the initial state S:

Walter's initial beliefs:	$\Box_{\mathbf{W}}^{\mathsf{root}_T}[\mathbf{M}, \succeq_{\mathbf{M}}]$,	$\Box_{\mathbf{W}}^{\mathsf{root}_T}[\mathbf{S}, \succeq_{\mathbf{S}*}]$,	$\Box_{\mathbf{W}}^{\mathsf{root}_T}\Box_{\mathbf{M}}^{\mathsf{root}_T}[\mathbf{S}, \succeq_{\mathbf{S}*}]$
Walter's beliefs at t_0:	$\Box_{\mathbf{W}}^{t_0}[\mathbf{M}, \succeq_{\mathbf{M}*}]$,	$\Box_{\mathbf{W}}^{t_0}[\mathbf{S}, \succeq_{\mathbf{S}*}]$,	$\Box_{\mathbf{W}}^{t_0}\Box_{\mathbf{M}}^{t_1}[\mathbf{S}, \succeq_{\mathbf{S}*}]$
Mary's beliefs:	$\Box_{\mathbf{M}}^{\mathsf{root}_T}[\mathbf{S}, \succeq_{\mathbf{S}}]$,	$\Box_{\mathbf{M}}^{\mathsf{root}_T}[\mathbf{W}, \succeq_{\mathbf{W}*}]$,	$\Box_{\mathbf{M}}^{t_0}[\mathbf{S}, \succeq_{\mathbf{S}}]$

The first row gives Walter's beliefs at root_T: he is right about Mary's preferences, but wrong about his own wife's preferences. He believes that Mary's shares his wrong belief about Sue. The second row gives Walter's beliefs at t_0: Here Walter revises his beliefs about Mary's preferences (for the detailed description, see below). Instead of changing his incorrect belief of Sue, he revises

his belief about Mary to $\succeq_{\mathbf{M}*}$ (which is wrong). The third row shows Mary's incorrect beliefs about Walter and her correct beliefs about Sue.

Of course, technically, S is not sufficiently specified with these data to compute the analysis of the game. We shall need statements about the Mary's beliefs of Walter's beliefs of Mary's beliefs. In the example, we assume that all beliefs other than the mentioned beliefs are trivial (*i.e.*, players believe that the other players are correctly informed about what they believe to be the state of affairs).

Let us give the analysis with our formalization from § 3 in words together with the subsequent computation of the nodes $\mathsf{t}^{[k]}_{T,S}$. We start at the root $\mathsf{t}^{[0]}_{T,S} := \mathsf{root}_T$. In root_T, Walter analyzes the tree from the point of view of what he believes are Mary's and Sue's beliefs. Based on the subjective states $S^t_{\mathbf{W}}$ (for $t \in T$), he can give the values of $\ell_S(\mathbf{W}, t)$ for all $t \in T$ by backward induction. We give the values of the function $t \mapsto \ell_S(\mathbf{W}, t)$ in Figure 2; we can read off the S-subjective run according to Walter by following the label of root_T to a terminal node: $\mathsf{t}^{[0]}_{T,\mathbf{W},S} = \mathsf{root}_T$ and $\mathsf{t}^{[1]}_{T,\mathbf{W},S} =$ ☼☼☼.

Note that this tree already incorporates the belief revision that happens in t_0 (even though it has no immediate effect): the values of the labelling at t_0 and all nodes below it depend on $S^{t_0}_{\mathbf{W}}$, not on $S^{\mathsf{root}_T}_{\mathbf{W}}$. Some of the formal properties of the function $t \mapsto \ell_S(\mathbf{W}, t)$ correspond to parts of the story: the fact that $\ell_S(\mathbf{W},$ ☼☼☼) corresponds to the sentence "*Walter convinced himself that Mary will never ask him to separate from Sue*", the fact that $\ell_S(\mathbf{W}, t_1) =$ ☼🐭☼ corresponds to "*... or –even worse– tell Sue about the affair.*"

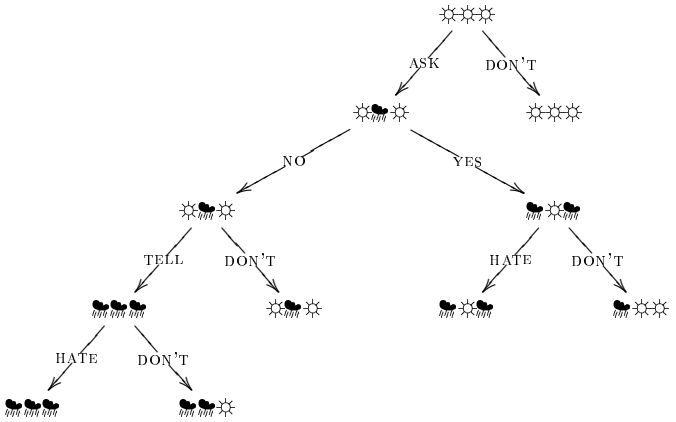


Fig. 2. Walter's subjective labelling: $t \mapsto \ell_S(\mathbf{W}, t)$.

Of course, Walter's subjective labelling at root_T is irrelevant for the game, as Mary has the first move. Mary's subjective labelling in root_T is given in Figure 3; again, we can read off Mary's (wrong) prediction of the outcome of the

game by following the label at the root to a terminal node, *i.e.*, $\mathsf{t}^{[0]}_{T,\mathbf{M},S} = \mathsf{root}_T$, $\mathsf{t}^{[1]}_{T,\mathbf{M},S} = t_0$, $\mathsf{t}^{[2]}_{T,\mathbf{M},S} = t_2$, and $\mathsf{t}^{[3]}_{T,\mathbf{M},S} =$. This is clearly represented in the storyline in "*[Mary] is convinced that if she presses Walter enough, he will finally leave Sue for her. She can make up with Sue afterwards.*"

In order to get the true first move, we have to compare the values of $\ell_S(\mathbf{M}, t_0) =$  and $\ell_S(\mathbf{M},$  $) =$ in Mary's true preference $\succeq_{\mathbf{M}}$. Since $\succeq_{\mathbf{M}}$, Mary will play ASK, and we get $\mathsf{t}^{[1]}_{T,S} = t_0$.

Fig. 3. Mary's subjective labelling: $t \mapsto \ell_S(\mathbf{M}, t)$.

The crucial element of the story is Walter's incorrect belief revision in t_0. Instead if revising his false beliefs about Sue, he revises his beliefs about Mary. In order to get our second true move, we consider the node t_0 in the labelling in Figure 2. We can read off Walter's prediction about what will happen if he plays YES or NO by following the labels of t_1 and t_2 to a terminal node. In Walter's true preference $\succeq_{\mathbf{W}}$, we compare $\ell_S(\mathbf{W}, t_1) =$  and $\ell_S(\mathbf{W}, t_2) =$. He chooses to play NO, and we get $\mathsf{t}^{[2]}_{T,S} := t_1$.

Now it is Mary's turn again. We consider t_1 in the labelling given in Figure 3 and compare $\ell_S(\mathbf{M}, t_3) =$  and $\ell_S(\mathbf{M},$ $) =$ via $\succeq_{\mathbf{M}}$ to get Mary's next move TELL and $\mathsf{t}^{[3]}_{T,S} := t_3$.

Finally, Sue just follows her true preference $\succeq_{\mathbf{S}}$ and we get $\mathsf{t}^{[4]}_{T,S} =$  .

With this formal analysis, we can make sense of counterfactual claims about the story. Let us give two examples for this:

Firstly, "If in t_0, Walter would have realized that he has misjudged his wife, he could have been better off". Let us assume that Walter revises his beliefs

correctly to $\Box^{t_0}_{\mathbf{W}}[\mathbf{S}, \succeq_{\mathbf{S}}]$ and $\Box^{t_0}_{\mathbf{W}}[\mathbf{M}, \succeq_{\mathbf{M}}]$, then he would have got the labelling given in Figure 4. In this situation, Walter would still have believed at root_T that ☼☼☼ is the true outcome, but then at t_0 would have (correctly) realized that he has the choice between 🌧☼☼ and 🌧🌧☼, triggering the move YES which leads to the outcome 🌧☼☼, preferred by both Walter and Mary over the actual outcome.

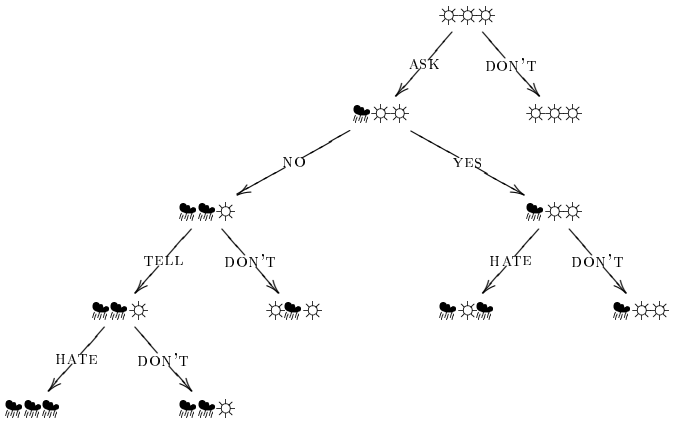


Fig. 4. First example of counterfactual reasoning: The alternative labelling for Walter with a correct belief revision.

As a second example, consider "Mary's false beliefs about Walter's preferences were irrelevant for the outcome". Suppose that $\Box^{\mathsf{root}_T}_{\mathbf{M}}[\mathbf{W}, \succeq_{\mathbf{W}}]$. In that case, Mary would have had the subjective labelling as described in Figure 5. Even though her prediction of the outcome changes drastically, her first move remains ASK, and thus the game will take exactly the same path.

5 Conclusion and Future Work

With the formal analysis of §§ 3 and 4, we fulfilled the goal mentioned at the end of § 2: we have a formal system that allows to mimic the intuitive reasoning of human beings about the game situation. However, the definition of a formal system provides only the very first step. A lot of open questions and problems remain.

Game-theoretic problems. Our analysis presupposes backward induction. This was not a problem in the given example, a game tree of depth 5. Of course, it is well-known that backwards induction is not a realistic assumption of human reasoning in interactive situations in general. It would be interesting to allow the players to choose other than the backwards induction strategies and add new

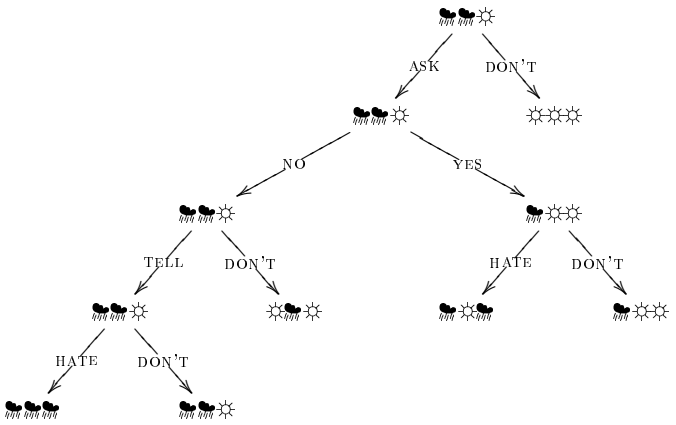


Fig. 5. Second example of counterfactual reasoning: The alternative labelling for Mary at root_T under the assumption of $\Box_{\mathbf{M}}^{\mathsf{root}_T}[\mathbf{W}, \succeq_{\mathbf{W}}]$.

atomic formulae $[i, \Sigma]$ interpreted as "player i follows the strategy denoted by Σ". In such a setting, we could investigate the analogue of Aumann's classic theorem [Au95] that common knowledge of rationality implies the backward induction solution. For this, we would have to give up the purely formal approach of § 3 and give definitions of rationality. First steps towards this goal are done in [LöPa$_0$Pe∞].

Complexity problems. Our formal system is anything but parsimonious: if $\langle I, T, \mu \rangle$ is an extensive game in which the tree has depth d, n terminal nodes and m non-terminal nodes, then there are $n!$ preferences, and thus $|I| \cdot n!$ many descriptions.

Only beliefs of the first d degrees are relevant for the analysis, so in order to count the states, we have to consider operator sequences of length $\leq d$. Hence there are at most

$$\left(\sum_{j=0}^{d} (m \cdot |I|)^j \right) \cdot |I| \cdot n! = \frac{(m \cdot |I|)^{d+1} - 1}{(m \cdot |I|) - 1} \cdot |I| \cdot n!$$

relevant states. No matter what you consider as input, this is a gigantic number hardly feasible for computations.

The example in § 4 gives a first clue as to what should be done here. In order to give a meaningful analysis of the story from § 2, we only needed six preferences (instead of $6! = 720$) and essentially nine elements of the state were enough to determine all relevant labellings.

In order to become useful, we would need to development techniques to systematically reduce the number of relevant entries in order to analyse a given situation.

Applications in logic. Discussions about beliefs in an interactive situation can quickly lead to a question about the existence[6] of a so-called universal belief space, *i.e.*, a space in which all possible beliefs are represented. Brandenburger and Keisler have discovered a fascinating paradox that suggests that it is not always possible to assume such a space exists [BrKe06]. Since we do not have negation in our language, Theorem 8.2 of [BrKe06] suggests that our framework is immune to the Brandenburger-Keisler paradox. Thus we hope that the underlying logic discussed in this paper may provide an answer to a question posed in [BrKe06]: find a logic $\mathcal{L}$ such that complete belief models exists for $\mathcal{L}$ (i.e., a complete model with respect to sets definable in $\mathcal{L}$) and the logic can be used to provide epistemic foundations of well-known solution concepts (such as the backward induction solution).

Other applications. We believe that a workable system using our formal analysis could have applications in applied artificial intelligence. First of all, it could allow artificial agents to mimic human behaviour in counterfactual reasoning about a given situation with actions and epistemic information that influences the actions.

As a second application, we could imagine reverse engineering of story lines (for instance, for computer games). The formal model may help humans to design a story line that is not too straightforward and involves an element of surprise, but at the same time is still understandable by the audience. The latter condition will require some complexity measure of the revisions involved and is closely connected to the mentioned complexity problems.

References

[Au95] Robert **Aumann**, Backward induction and common knowledge of rationality, **Games and Economic Behavior** 8 (1995), p. 6-19

[Au99] Robert **Aumann**, Interactive epistemology I: knowledge, **International Journal of Game Theory** 28 (1999), p. 263-300

[BaSi99] Pierpaolo **Battigalli**, Marciano **Siniscalchi**, Hierarchies of conditional beliefs and interactive epistemology in dynamic games, **Journal of Economic Theory** 88 (1999), p. 188-230

[Bo$_0$04] Oliver **Board**, Dynamic interactive epistemology, **Games and Economic Behavior** 49 (2004), p. 49-80

[Bo$_1$Ba99] Giacomo **Bonanno**, Pierpaolo **Battigalli**, Recent results on belief, knowledge and the epistemic foundations of game theory, **Research in Economics** 53 (1999), p. 149-225

[Br06] Adam **Brandenburger**, The power of paradox: some recent developments in interactive epistemology, *to appear in:* **International Journal of Game Theory** 2006

[6] This is not just a theoretical question — both [BaSi99] and [BrKe01] give an analysis of solutions concepts which rely on the existence of a "universal belief space". See [Br06] for a discussion.

[BrKe01] Adam **Brandenburger**, H. Jerome **Keisler**, Epistemic conditions for iterated admissibility, *in:* Johan van Benthem (*ed.*), Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2001), Certosa di Pontignano, University of Siena, Italy, July 8-10, 2001, Morgan Kaufman 2001, p. 31-37

[BrKe06] Adam **Brandenburger**, H. Jerome **Keisler**, An impossibility theorem on belies in games, *to appear in:* **Studia Logica**

[dB04] Boudewijn **de Bruin**, Explaining Games: On the Logic of Game Theoretic Explanations, PhD Thesis, Universiteit van Amsterdam 2004, **ILLC Publications** DS-2004-03

[Fa+95] Ronald **Fagin**, Joseph Y. **Halpern**, Yoram **Moses**, Moshe Y. **Vardi**, Reasoning about Knowledge, MIT Press, 1995

[Gä92] Peter **Gärdenfors**, Belief revision: an introduction, *in:* Peter Gärdenfors (*ed.*), Belief Revision, Cambridge University Press 1992 [Cambridge Tracts in Theoretical Computer Science 29], p. 1-28

[Lö06] Benedikt **Löwe**, Revision Forever!, *to appear in:* Henrik Schärfe, Pascal Hitzler, Peter Øhrstrøm (*eds.*), Proceedings of the 14th International Conference on Conceptual Structures, ICCS06, Aalborg, Denmark, 2006, Springer-Verlag, Berlin, 2006 [Lecture Notes in Artificial Intelligence]

[LöPa$_0$Pe∞] Benedikt **Löwe**, Eric **Pacuit**, Andrès **Perea**, A syntactic approach to beliefs in games, *in preparation*

[Mo+97] Phlippe **Mongin**, Michael O. L. **Bacharach**, Louis-André **Gérard-Varet**, Hyun Song **Shin** (*eds.*), Epistemic logic and the theory of games and decisions, Kluwer Academic Publishers, 1997 [Theory and Decision Library, Series C: Game Theory, Mathematical Programming and Mathematical Economics 20]

[OsRu94] Martin J. **Osborne**, Ariel **Rubinstein**, A Course in Game Theory, MIT Press, 1994.

[Pa$_1$03] Rohit **Parikh**, Levels of knowledge, games, and group action, **Research in Economics** 57 (2003), p. 267-281

[Pe05] Andrès **Perea**, A model of minimal probabilistic belief revision, *preprint* 2005

[St98] Robert **Stalnaker**, Belief revision in games: forward and backward induction, **Mathematical Social Sciences**, 36 (1998), p. 31-56