Modal Models for Games with Unawareness

Michael Franke
Department of Linguistics
University of Tübingen

July 18, 2010

Abstract

In games with unawareness players may entertain fairly different representations of their strategic choice situation, and may also reason about these possibly different subjective conceptualizations of the game they are playing. This paper uses basic tools from modal logic, especially modal model theory, to represent games with unawareness of moves and to assess the implications of rationality and various forms of belief therein for this class of games.

1 Introduction

Traditional game theory assumes that a game model captures a strategic choice situation the way that it actually is (or rather, how it appears to the modeller). Traditional game theory also assumes that this actual (or modeller's) view coincides exactly with how the game is conceived by all players at each of their choice points. This assumption is often summarized plainly as requiring that the game be common knowledge among the players. But, clearly, this strong assumption is not always warranted. In real life we may often be engaged in strategic choice situations without reven representing them to us as such, or participants of a strategic choice situation may have rather different ideas of what options are available and what parameters are relevant to the interactive choice.

In order to accommodate for such diverging *subjective conceptualizations* of a game, a modest number of models for these situations have surfaced recently, giving evidence to the rising interest in the epistemic foundations of game theory, not only as far as the solution concepts, but also as far as the game models are concerned (e.g. Feinberg 2004; Halpern and Rêgo 2006; Ozbay 2007; Feinberg 2009; Heifetz et al. 2010). Following suit, this paper studies *games with unawareness of moves*, i.e., games in which players may be unaware that some relevant choice or parameter exists, and in which players may have various beliefs about this limited awareness of others. (Section 5.1 discusses a possible extension to games with unawareness from indistinguishability, i.e., games where players may be unable to distinguish certain moves or states of affairs.)

The relevant notion of unawareness stems from formal epistemology, situated at the interface between theoretical computer science, logic and philosophy. It has first been introduced by Fagin & Halpern (Fagin and Halpern 1988) as a means of solving issues with, foremost, *logical omniscience* (the problem that in standard model logics of belief and knowledge agents know all tautologies and consequences of their knowledge). An agent is unaware of a contingency if, intuitively speaking, she has no mental representation of it and consequently lacks all explicit beliefs about it. There are different sources of such unawareness, not all of which will have equal properties under all circumstances (de Jager 2009). For instance, agents may be unaware because they lack certain concepts (e.g., a novice chess player yet unfamiliar with castling), or they may be unaware because they fail to pay due attention (e.g., a novice chess player familiar with castling who has temporarily forgotten). The relevant property that concerns us here is that an agent who is unaware cannot by mere introspection alone level her unawareness. (This also assumes that agents are not aware of their own unawareness. To give up this assumption raises many interesting issues, to which we tend briefly in Section 5.2.)

If an agent is partially unaware, then this will also restrict the possible beliefs she may have about what others believe. An agent who is partially unaware cannot explicitly ascribe beliefs to other agents involving contingencies that the agent herself is unaware of. Glossing crudely over interesting formal detail here, this is part of the reason why unawareness is not sufficiently captured as a variety of standard belief (cf. Dekel et al. 1998), and why therefore formal models of interactive awareness must constrain the space of an agent's possible explicit beliefs either *syntactically* by defining subjective languages exclusively in terms of which each agent can represent to herself any of her ordinary or higher-oder beliefs (cf. Fagin and Halpern 1988; Halpern 2001), or *semantically* by imposing a layered structure on the state space or space of possible worlds in which the contingencies epistemically accessible to each agent are located (Modica and Rustichini 1999; Heifetz et al. 2006, 2008; Halpern and Rêgo 2008; Li 2009). (The paper of Halpern & Rêgo Halpern and Rêgo 2008 provides a good overview and comparison of models.)

In line with this, games with unawareness of moves are games with certain *restrictions* on players' beliefs about (players' beliefs about ...) the structure of the game. Also intuitively, the possibility of unaware players has strategic implications on rational behavior in games: making others aware may be beneficial (e.g., making consumers aware of a new product) or extremely unwanted (e.g., making a spouse aware of even the possibility of an affair). Indeed, restrictions on the players' (beliefs about) awareness of moves severely impact, in fact weaken, the implications of rationality. An unaware player chooses rational, if she maximizes her gain in expectation *given her subjectively limited representation* of the choice situation. This also means that only in marginal cases would these interactive belief restrictions of a game with unawareness be amenable to treatment under standard games with imperfect information and their standard solution concepts.

In keeping with the logical tradition of awareness, this paper offers a manageable alternative representation of games with unawareness in terms of *awareness structures*, which are essentially pointed models for a basic modal language of belief with few additional constraints. Awareness structures capture succinctly which parts of the game each agent is aware of, and what every player's higher-order beliefs about other player's awareness are. These game models thereby provide a scaffolding for the possible beliefs that players can have about the way the game is played by rational agents. It is indeed straightforward to extend these models to full-fledged epistemic models in the vein of Robert Stalnaker (Stalnaker 1994, 1998) in which the implications of (belief in) rationality can be studied. The focus of this paper is in highlighting the benefits of linking game theory with modal logic. Established results from modal model theory facilitate the representation, comparison and handling of games with unawareness and their epistemic models.

The paper is structured as follows. Section 2 introduces the notion of a partial representation of a dynamic game obtained from pruning the game tree. This will be the main source of unawareness that is captured in the awareness structures introduced in Section 3. This section shows how awareness structures are versatile hybrids, game representations *and* models for a basic modal language to describe awareness states and beliefs about these. Section 4 extends awareness structures to full epistemic models of games with unawareness and discusses briefly some of the implications of rationality and belief therein. Section 5 concludes speculatively with a number of possible extensions of the modal approach taken here.

2 Dynamic Games and Pruning

There are several possibilities of defining a dynamic game. For the present purposes, it helps to look at a dynamic game as a tree with ornamentation (cf. Kreps and Wilson 1982): the tree defines the basic structure of the play, and the ornamentation interprets the structure by assigning players to choice nodes, action labels to choices, utilities to the leaves of the tree etc. The reason why this representation proves helpful in the present context is because it makes it easy to think about *partial representations* of a game G as being obtained from G by restricting the game tree while preserving the labeling. Also, if a partial representation contains parts of the original game tree, it is easy to identify choice points across two games as being identical.

2.1 Dynamic Games as Ornamented Trees

Definition 2.1 (Dynamic Game with Imperfect Information). A dynamic game with imperfect information is a structure

$$G = \langle H, <, N, \{A_i\}_{i \in N}, P, A, \{u_i\}_{i \in N}, \Pr, \{V_i\}_{i \in N} \rangle$$

that consists of (i) a game tree, (ii) a collection of labels, and (iii) a labeling that assigns labels to elements of the game tree. In particular:

- $\langle H, < \rangle$ is a *game tree*:
  - H is a (finite) set of *histories*, or decision nodes
  - $<$ is a partial order on H such that:
    - there is a unique $<$-minimal element h_0, called the *root*

        - every history $h \neq h_0$ has exactly one *predecessor*, namely the unique $<$-maximal element of the set $\{h' \in H \mid h' < h\}$
    - elements in $Z = \{h \in H \mid \neg\exists h' \in H\ h < h'\}$ are called *terminal histories*
    - every non-terminal history h has a non-empty set of *successors* $H(h)$, defined as the set of all $<$-minimal elements in $\{h' \in H \mid h < h'\}$
- $\langle N, \{A_i\}_{i \in N}\rangle$ are *labels*:
    - $N = \{1, 2, \ldots n\}$ is a the set of players with designated player n as Nature
    - A_i is a set of actions for player i
- $\langle P, A, \{u_i\}_{i\in N}, \Pr, \{V_i\}_{i \in N}\rangle$ is the *labeling*:
    - $P : H \setminus Z \to N$ is a *player function*
    - $A : H \times H \to \bigcup_{i \in N} A_i$ assigns action labels to each choice as follows:
        - $A(h, h')$ is only defined if $h' \in H(h)$
        - $A(h, \cdot) \mapsto A_{P(h)}$ is an injection
    - $u_i : Z \to \mathbb{R}$ is a *utility function* for all $i \in N$, $i < n$
        - we write $u_i(z)$ for the i-th component of $u(z)$
    - $\Pr$ is a function that gives for each nature move $h \in P^{-1}(n)$ a probability distribution $\Pr_h \in \Delta(H(h))$ over successors of h
    - $V_i \subseteq \mathcal{P}(H \setminus Z)$ is a set of information states of player i such that:
        - $\bigcup_{i \in N} V_i$ is a partition of $H \setminus Z$
        - if $v \in V_i$ and $h, h' \in v$, then $P(h) = i$ and, moreover, $A(h, \cdot)$ and $A(h', \cdot)$ have the same image sets
        - notice that V_n is a set of singletons and can be treated as $\bigcup V_n$.

Observe that with this definition a game may have *redundant labels*: it may list players in N that never move according to P, or if it may contain some action of some player i in A_i that is never assigned to any of i's decision nodes by A. In the following, attention is restricted to games with non-redundant labels.

Let $\mathfrak{G}$ be the collection of all dynamic games with imperfect information.

Say that an action $a_i \in A_i$ is *available* to player i at choice point v if there is an $h \in v$ and an $h' \in H(h)$ such that $a_i = A(h, h')$. A *pure strategy* s_i for player i is a function from V_i to actions available at each $v \in V_i$. A *behavioral strategy* σ_i for player i is a function from V_i to a probability distribution over actions available to i at each $v \in V_i$. A pure (behavioral) strategy profile for game G is a tuple $s = \langle s_1, \ldots, s_{n-1}\rangle$ ($\sigma = \langle \sigma_1, \ldots, \sigma_{n-1}\rangle$) of pure (behavioral) strategies for each player.

Notice that we have not constrained the above definition to ensure *perfect recall* of all agents. Nothing of substance hinges on this omission. Rather, it is simply convenient for the central concerns of this paper not to stick to the perspective of a single player, but rather to the set $V = \bigcup_{i<n} V_i$ of all choice points of non-nature players. Additional constraints on the temporal coherence of an agent's beliefs and rational choices (also in the light of unawareness) can be added if necessary. For the main purposes of this paper —a concise representation of games with unawareness— none of this is necessary. In line with this, a pure (behavioral) strategy profile will occasionally be characterized as function from V to (probability distributions over) actions available in each v, where this is feasible.

2.2 Partial Representations from Pruning

Representing a game G as ornamented trees makes it easy to define a *partial representation* of G as obtained by either (i) pruning its game tree, or (ii) lumping together a number of branches as subjectively indiscernible. These partial representations then serve as a means of representing agents' unawareness of the game. An agent who represents the actual game as a pruned version of it is unaware of the pruned actions and states of affairs. An agent who represents the actual game as a lumped version of it cannot distinguish between the lumped actions and states of affairs. We concentrate on pruning, as this affords less notational effort, and discuss an extension to lumping in Section 5.1.

Intuitively speaking, G' is a *pruning* of G, $G' \sqsubseteq G$, if G' is constructed from G by removal of nodes from the game tree of G while all the labeling is preserved as far as possible. If we make sure that no non-terminal nodes in G become terminal nodes, then the definition of pruning is entirely straightforward.

Definition 2.2 (Pruning). Take two dynamic games with imperfect information (and non-redundant labeling):

$$\begin{aligned} G &= \langle H, <, N, \{A_i\}_{i\in N}, A, P, \{u_i\}_{i\in N}, \Pr, \{V_i\}_{i\in N}\rangle \\ G' &= \left\langle H', <', N', \{A'_i\}_{i\in N'}, A', P', \{u'_i\}_{i\in N}, \Pr', \{V'_i\}_{i\in N'} \right\rangle . \end{aligned}$$

G' is a *pruning* of G, $G' \sqsubseteq G$, if the following conditions hold:

- $H' \subseteq H$ (prune the game tree)
- $<' = < \upharpoonright H'$ (restrict order to elements left)
  (define Z' and $H'(h)$ in the obvious ways)
- $Z' \subseteq Z$ (no new terminal nodes)
- $N' \subseteq N$, and $A'_i \subseteq A_i$ (preserve non-redundant labels)
- $A'(h, h') = A(h^*)$, where h^* is the unique element such that $h^* \in H(h)$ and $h^* \leq h'$
  (preserve labels, even if successors are pruned)
- $P' = P\upharpoonright(H' \setminus Z')$ (restrict to remaining elements)
- $u'_i = u_i\upharpoonright Z'$
- $\Pr'_h = \Pr_h(\cdot|H'(h))$ for all $h \in P'^{-1}(n)$ (restrict to remaining nature moves)
- $V'_i = \{v \cap H' \mid v \in V_i\} \setminus \{\emptyset\}$. (restrict each information states)

The set of all pruned games of G corresponds one-to-one with the set of all subsets $H' \subseteq H$ such that $Z' \subseteq Z$. With this it is immediate that the relation $\sqsubseteq$ gives rise to a partial order on games, being reflexive, anti-symmetric, and transitive. Moreover, it is clear that whenever $G' \sqsubseteq G$ we can easily identify choice nodes and information sets across games. Say that a node $h \in H$ *reoccurs* in G' iff $h \in H'$. Similarly, for information states of G, consider the set $V = \bigcup_{i<n} V_i$ of all choice points of non-nature players as the set of *(generally relevant) subjective views*. Say that a view v from G *reoccurs* in G' iff some $h \in v$ reoccurs in Γ'. Abusing notation, write $h \in G'$ and $v \in G'$ whenever h or v reoccur in G'. For any two games $G' \sqsubseteq G$, since pruning leaves information states otherwise unaltered, if v is an information state from G with $v \in G'$, v can be used to refer to a *unique* information state in G' even if not all elements of v reoccur in G'. This kind of trans-game identification, as we could call it, is crucial for the way that we represent games with unawareness in terms of awareness structures.

3 Awareness Structures

A game with unawareness should be an enriched representation, based on a classical dynamic game G, that specifies how G occurs to each subjective view $v \in V = \bigcup_{i<n} V_i$, i.e., all information states of non-nature players in G, together with how each view v_1 believes that each view v_2 believes that $(\dots)$ G occurs to v_n. Formally, there are several ways of representing any such an infinite set of intersubjective views. One possibility is to define a game with unawareness as a collection of classical games with a function that maps each choice point of each game to some choice point in some possibly different game in the collection (Halpern and Rêgo 2006; Heifetz et al. 2010). Another possibility is to define a game with unawareness as a collection of games, one for each (relevant) finite sequence of views $v_1 v_2 \dots v_n$ (Feinberg 2009).

This paper suggest a further possibility that has several methodological advantages that will become clear as we go along. I propose to represent games with unawareness in terms of *awareness structures*, which are essentially a special kind of (interpreted and pointed) relational Kripke-structure, as familiar from modal logic.

Definition 3.1 (Awareness Structure). Let G be a dynamic game with imperfect information with information states $V = \bigcup_{i<n} V_i$ of all non-nature players. An awareness structure based on G is a tuple $\mathcal{A}_G = \langle W, w_0, \{R_v\}_{v\in V}, L\rangle$ such that:

- W is a set of possible worlds,
- w_0 is the actual world (specifying the modeller's view),
- $R_v \subseteq W \times W$ is an accessibility relation for the viewpoint $v \in V$,
- $L : W \to \mathfrak{G}$ assigns to each world w a game $L(w)$.

Moreover, $\mathcal{A}_G$ needs to satisfy the following constraints:

Centering: $L(w_0) = G$,

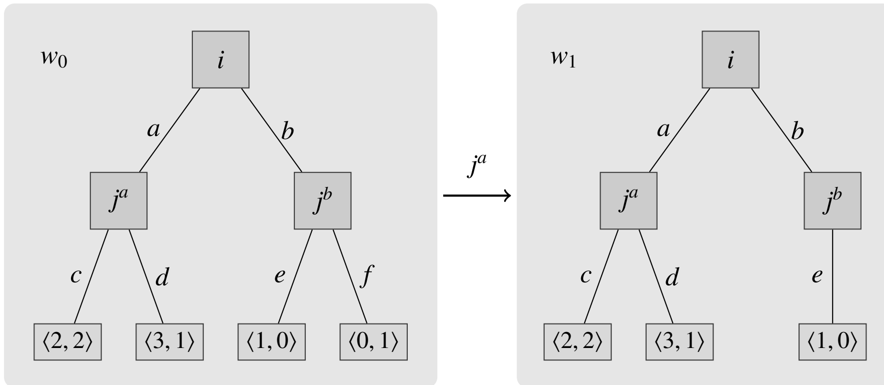

Figure 1: A simple awareness structure

REDUCTION: if wR_vw', then $L(w') \sqsubseteq L(w)$,

EXISTENCE: if v is an information state in game $L(w)$, then there is a world w' such that wR_vw',

RELEVANCE: whenever wR_vw' then v is an information state in $L(w)$ that reoccurs in $L(w')$,

INTROSPECTION: for all v the relations R_v are transitive and Euclidean.[1]

Given awareness structure $\mathcal{A}_G$, define the set of worlds accessible from w via a sequence $v_1v_2\dots v_n$ of views from $\mathcal{V}$ as:

$$R_{v_1}(w) = \left\{w' \in W \mid wR_{v_1}w'\right\}$$
$$R_{v_1\dots v_{n-1}v_n}(w) = \bigcup_{w' \in R_{v_1\dots v_{n-1}}(w)} R_{v_n}(w')\,.$$

Also define accessibility relation R^* as accessibility via some chain of R_v-s:

$$R^*(w) = \left\{w' \in W \mid \exists v_1 \dots v_n \ : \ w' \in R_{v_1\dots v_n}(w)\right\}\,.$$

For convenience, we will restrict attention to awareness structures which are *connected* in the sense that $W = R^*(w_0)$.[2]

In awareness structure $\mathcal{A}_G = \langle W, w_0, \{R_v\}_{v\in V}\,, L\rangle$, each world $w \in W$ is associated with a game $L(w)$ by interpretation function L. By reduction, connectedness and transitivity of $\sqsubseteq$, any $L(w)$ is a *partial representation* of G obtained from pruning the game tree. We look at accessibility relations $R_v \subseteq W \times W$ for each information state v that occurs in the actual game G, because these are the views that we have to assign a subjective representation of G to. Notice that it is possible to identify information states from G uniquely in all its prunings by reoccurrence. The existence and relevance constraints make sure that all and only choice points in $L(w)$ are associated with a recognizable choice point in some game $L(w')$. More concretely, the existence constraint ensures that all choice points in game $L(w)$ are associated with some subjective view. The relevance constraint ensures that (i) only occurring choice points associate with a subjective view of the game, and that (ii) whenever wR_vw' it is clear which choice point in $L(w')$ we are relating v to, namely the unique subset of v that is an information state in $L(w)$. Finally, the introspection constraint forces the proper behavior of nested beliefs an agent may have about her *own* conceptualization of the game.

Here is a simple example. The awareness structure in Figure 1 has only two possible worlds $W = \{w_0, w_1\}$. The game $L(w_0)$ is actual: first it is player i's choice between actions a and b, then player j chooses between c and d, or e and f. However, this is not how all players conceptualize the game at all choice points. Subjective views are given by the accessibility relations, where all necessary reflexive arrows are omitted for readability. Thus, according to the awareness structure in Figure 1, player i is aware of the game as it is, and so is player j after observing b. However, player j is not aware of her action f after observing a: in game $L(w_1)$, which is how j^a construes the game in the actual world w_0, the action f is pruned away.

[1] A relation $R \subseteq W \times W$ is Euclidean if w_1Rw_2 and w_1Rw_3 implies w_2Rw_3.

[2] Worlds that are not connected to actual world w_0 in this sense are irrelevant for the interpretation of modal expressions with a "local" semantics, and we will only be interested in those.

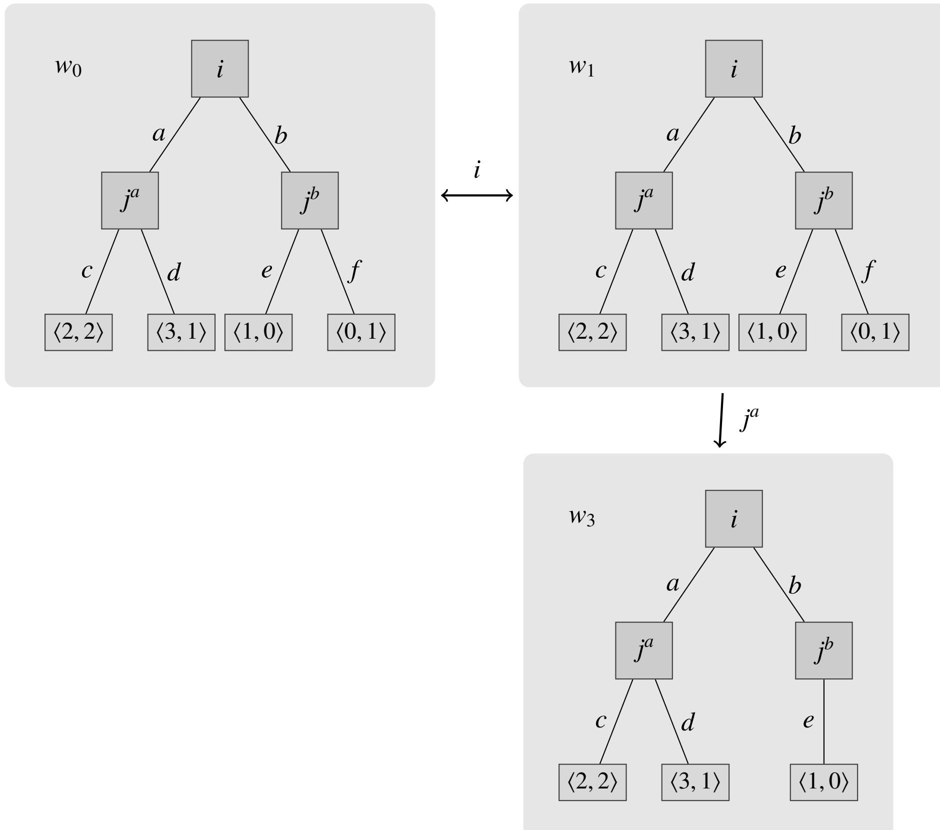

Figure 2: Awareness structure that captures uncertainty of awareness

In summary, the awareness structure in Figure 1 encodes exactly which nodes of the actual game each player at each choice point is aware of. But this is not all that is represented in this awareness structure. There is also higher-order information, so to speak, about, for instance, how player i conceptualizes how j^a conceptualizes the game. Using terminology I will justify below, we could say that player i *knows* that j^a is unaware of f, but also that, for instance, i knows that j^b knows that j^a is unaware of f. However, j^a does not know that he is unaware, as the only world she has access to is w_1.

Coincidentally, all of the higher-order information represented in Figure 1 does not include any uncertainty, but that is not necessarily the case for all awareness structures, as Figure 2 demonstrates. Here player i doesn't know whether j^a is aware of f, but j^b *does* know that j^a actually is. And, both j^a and j^b know that i is uncertain about j^a's awareness state.

These examples and the locutions I have used to describe the representations suggest that awareness structures are *a representation of beliefs* of the players that are relevant for reasoning about awareness of moves of the game. Indeed, formally speaking, the constraints on awareness structures closely ressemble a basic modal logic of belief (c.f. Halpern 2003). Normally, modal logics for belief feature accessibility relations that are serial, transitive and Euclidean.[3] The latter two are taken over immediately, and existence plays the role of *seriality* in awareness structures. On top of that, however, awareness structures have additional constraints. We may think of the reduction constraint as essentially imposing a lattice structure on the state space W to which agents' beliefs are susceptible, much like in logics of interactive unawareness (e.g. Heifetz et al. 2006, 2008; Li 2009). So, what kind of beliefs are represented exactly in awareness structures? To answer this question, we will define a simple modal language to talk about games with unawareness, and use awareness structures as formal models for this modal language.

[3] A relation $R \subseteq W \times W$ is serial if for all $w_1 \in W$ there is a $w_2 \in W$ such that $w_1 R w_2$.

3.1 Modal Language of Games with Unawareness

Fix a game G with nodes H and look at the simple modal language $\mathcal{L}^G$ given by

$$\varphi := \; h \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond_v(\varphi)\,.$$

In words, $\mathcal{L}^G$ is the smallest set of formulas that contains all $h \in H$ and that is closed under negation, conjunction and the modal operators $\Diamond_v$ for each view $v \in V$ from G. We define disjunction $\vee$, material implication $\rightarrow$ and material equivalence $\leftrightarrow$ in the usual way. Also, we define the modal operator $\Box_v$ as the dual of $\Diamond_v$, as usual: $\Box_v\varphi = \neg\Diamond_v\neg\varphi$.

The language $\mathcal{L}^G$ is a basic modal language based on a set of proposition letters made up of H and awareness structures are straightforward models for this basic language. Fix awareness structure $\mathcal{A}_G = \langle W, w_0, \{R_v\}_{v\in V}, L\rangle$ and define the following basic semantics for expressions in $\mathcal{L}^G$:

$$\begin{aligned}
\mathcal{A}, w \models h \quad &\text{iff} \quad h \in L(w)\\
\mathcal{A}, w \models \neg\varphi \quad &\text{iff} \quad \mathcal{A}, w \not\models \varphi\\
\mathcal{A}, w \models \varphi \wedge \psi \quad &\text{iff} \quad \mathcal{A}, w \models \varphi \text{ and } \mathcal{A}, w \models \psi\\
\mathcal{A}, w \models \Box_v(\varphi) \quad &\text{iff} \quad \mathcal{A}, w' \models \varphi \text{ for all } w' \in R_v(w)\,.
\end{aligned}$$

If proposition h is true in world w we should interpret this as saying that node h occurs in the game that would be actual if w was the actual world. Formula $\Diamond_v\varphi$ expresses that the player at choice point v considers it possible that φ, and, similarly, $\Box_v\varphi$ expresses that the player at v believes that φ.

Using standard definitions, we can also express what it means for an agent at v to know something and when a proposition is common belief or common knowledge. We say that the agent at v knows φ if φ is true and the agent at v believes it. The common belief operator cannot be defined in terms of expressions from $\mathcal{L}^G$ (because we would need an *infinite* conjunction of simple belief formulas) but it may be added when convenient:

$$\mathcal{A}, w \models \mathrm{CB}\varphi \quad \text{iff} \quad \mathcal{A}, w' \models \varphi \text{ for all } w' \in R^*(w)$$

Finally, for the purposes of this paper, define common knowledge simply as true common belief.

Most importantly, we can express that the player at v is aware of node h by forumla $\Diamond_v h$. It will become transparent in the following that the language $\mathcal{L}^G$, if interpreted on awareness structure $\mathcal{A}_G$, really expresses two things: (i) what each agent at each viewpoint would be aware of in each world, if that world was actual, and (ii) what each agent at each viewpoint would believe in each world, if that world was actual, about what everybody else (believes about what everybody else ...) is aware of.

Of course, awareness structures are not just unrestricted models for $\mathcal{L}^G$. The definition of awareness structures imposes certain constraints on the accessibility relations and the games that can be associated with each world. But what exactly is the relation between awareness structures and normal modal models for the language $\mathcal{L}^G$? Clearly, under the above semantics we can think of function $L(w)$ in awareness structures as a valuation function, based on the set of proposition letters H. By construction of pruning, suitable subsets $H' \subseteq H$ will be associated with exactly one pruned version of the underlying G. This is why, effectively, an awareness structure *is* just a special kind of *pointed modal model*, i.e., a model with a designated actual world w_0 that is furthermore guaranteed to make certain formulas true. This is the content of the following proposition whose proof is in the appendix.[4]

Proposition 3.2. A pointed modal model $M = \langle W, w_0, \{R_v\}_{v\in V}, L\rangle$ for language $\mathcal{L}^G$ is an awareness structure based on G iff, based on the standard semantics, for all $h \in H$, $v \in V$ and $w \in W$ the following statements are all true:

No New Terminal Nodes: $M, w \models h \rightarrow z(h)$

Centering: $M, w_0 \models h$

Reduction: $M, w \models \Diamond_v h \rightarrow h$

Existence: $M, w \models v^* \rightarrow \Diamond_v \top$

Relevance: $M, w \models \Diamond_v \top \rightarrow (v^* \wedge \Diamond_v v^*)$

Introspection: $M, w \models \Box_v\varphi \rightarrow \Box_v\Box_v\varphi$ and $M, w \models \neg\Box_v\varphi \rightarrow \Box_v\neg\Box_v\varphi\,.$

[4] As for notation, let $z(h)$ be the disjunction $h_1 \vee \cdots \vee h_n$ for all terminal successors of h, let $v^* = (h_1 \vee h_2 \vee \cdots \vee h_n)$ be the disjunction of all choice nodes in information state v, and let $\top$ denote "top", a forumla that is true in all possible worlds.

These constraints on modal models guarantee that further formulas from $\mathcal{L}^G$ are valid on the class of awareness structures, i.e., that theses formulas hold of all awareness structures $\mathcal{A}$ and worlds w that occur in $\mathcal{A}$. If formula φ is valid in this sense, we write $\models \varphi$. The most surprising validity is certainly that, where relevant, the meaning of modal operators $\Diamond_v$ and $\Box_v$ coincide for literals h and $\neg h$. This is crucial in order to understand what awareness structures model precisely. So, first fix what "relevant" means here, by defining the set of *belief types* that occur in $\mathcal{A}_G$ as the set $B_{\mathcal{A}_G} = \{\langle w, v\rangle \mid w \in W \wedge v \in L(w)\}$.

Proposition 3.3. For all belief types $\langle w, v\rangle \in B_{\mathcal{A}_G}$ and all $h \in H$: $\mathcal{A}, w \models \ \Diamond_v h \ \leftrightarrow \ \Box_v h$.

A proof of this is easy, once we acknowledge a generally noteworthy fact about awareness structures.

Fact 3.4. Given awareness structure $\mathcal{A}_G$, each belief type $\langle w, v\rangle$ is associated with exactly one game, for which we write $L(w, v)$.

Proof of Fact 3.4. We need to show that for all w and v, for all $w_1, w_2 \in R_v(w)$ we have $L(w_1) = L(w_2)$. This is a direct consequence of the conjunction that R_v is Euclidean and that $\sqsubseteq$ is anti-symmetric. □

Proof of Proposition 3.3. That $\mathcal{A}, w \models \Diamond_v h \to \Box_v h$ follows from Fact 3.4 and that $\mathcal{A}, w \models \Box_v h \to \Diamond_v h$ for occurring belief types is also obvious, because if $\langle w, v\rangle \in B_{\mathcal{A}_G}$, then by the existence constraint $R_v(w)$ is non-empty, and so if all $w' \in R_v(w)$ make h true, then there is a world $w' \in R_v(w)$ that makes h true. □

Proposition 3.3 states a possibly surprising, but crucial result on the expressivity of awareness structures. To rephrase it once more, it says that, where relevant, certain distinctions that *could* be expressed by $\mathcal{L}^G$ on normal modal models *cannot* be expressed by $\mathcal{L}^G$ on the restricted set of modal models that satisfy the constraints of awareness structures. What this means is that $\mathcal{L}^G$, if evaluated on awareness structures, can only express exactly *one* first-order propositional attitude: whether an agent is aware of h or not. Call these statements *awareness statements*. With Proposition 3.3 and the definition of modals, it is clear that we can express the awareness statement that "the player at v is aware of h" by any of the formulas $\Diamond_v h$, $\Box_v h$, $\neg\Diamond_v\neg h$ or $\neg\Box\neg_v h$. Similarly, we can express the awareness statement that "the player at v is not aware of h" by any of the formulas $\Diamond_v\neg h$, $\Box_v\neg h$, $\neg\Diamond_v h$ or $\neg\Box_v h$. Although this is a restriction on expressivity due to the constraints on models, this is how it should be if we model awareness of moves: an agent is aware of a node if and only if she believes it is part of the relevant structural description of the game. Plain uncertainty about nodes is different in kind and can be modelled within the confines of classical game models if necessary.

Of course, an equivalent of Proposition 3.3 does not hold for arbitrary formulas. For arbitrary φ, only the trivial statement holds that for all occurring belief types $\langle w, v\rangle \in B_{\mathcal{A}_G}$ we have:

$$\mathcal{A}, w \ \models \ \Box_v(\varphi) \ \to \ \Diamond_v(\varphi)\,.$$

The reverse is generally not the case. This means that $\mathcal{L}^G$, evaluated on awareness structures, captures quite the usual interactive beliefs about which awareness statements are true. In summary, we find that awareness structures can express two things: (i) all relevant awareness statements, and (ii) all relevant beliefs about (beliefs about ...) awareness statements.

What is left to argue is that this is exactly the right portion of expressivity that we need in order to represent games with unawareness of moves. To motivate this, consider another example. Take the game G_0 in Figure 3, a simple two player extensive form game, with nodes $H = \{h_0, \ldots, h_4\}$, two players i and j with one choice each, and labels and outcomes as given in the diagram. What is a modeller interested in, when describing a game with unawareness of moves based on this game? In a classical game, the game structure is common knowledge. That means that all awareness statements $\Diamond_v h$ are common knowledge. In specifying, even in thinking about a game with unawareness, we register all deviations from this rule. For instance, in a particular case we might be interested in modeling a situation where both players are aware of the game as it is, i.e., aware of all nodes, while player j knows that player i believes that player j is unaware of action c, i.e., unaware of node h_3. This modeling intention is expressible by a conjunction of (i) truth and falsity of all awareness statements and (ii) all beliefs about (i) and, recursively, (ii). This is expressible formally in the language $\mathcal{L}^{G_0}$. For the current example, the most relevant statements are:[5]

1. $\Diamond_i h \wedge \Diamond_j h$ for all $h \in H$
2. $\mathrm{C}(\Diamond_i h \wedge \Diamond_j h)$ for all $h \in \{h_1, h_2\}$

[5]Strictly speaking, the common belief operator is not in $\mathcal{L}^G$, but that is insubstantial as we may replace it with the equivalent infinite conjunction of regular belief statements.

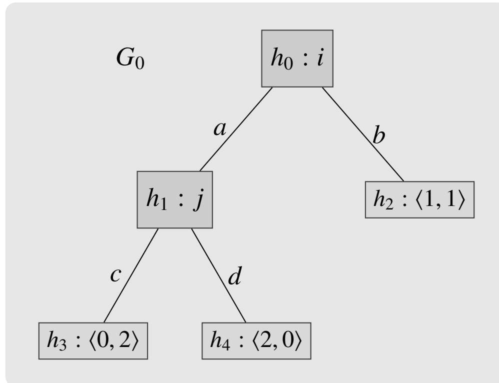

Figure 3: Simple Dynamic Game

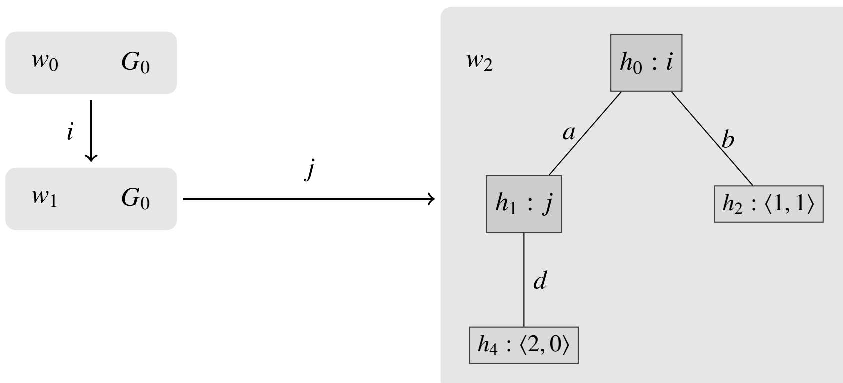

Figure 4: Awareness structure based on game G from Figure 3

3. $\Box_i \neg \Diamond_j h_3$
4. $\Box_j \Box_i \neg \Diamond_j h_3$

If our modeling intention is consistent —we surely find out when it is not— we find an awareness structure that models the formal description of our modeling intention, for instance, in the one given in Figure 4, where there are three possible worlds $W = \{w_0, w_1, w_2\}$ where both w_0 and w_1 are associated with game G_0 from Figure 3. It is easy to verify that this awareness structure makes all the formulas above true, and it is in this sense that it is a proper representation of the game with unawareness of moves that we had wanted to model.

The example illustrates two things. Firstly, it highlights the close connection between, on the one hand, the intention what to model and a language to express that intention in and, on the other hand, a formal language and its models to make all this precise. As awareness structures play a dual role, as either game models or modal models for a language that describes our modeling intentions, these structures mitigate between modeling intention and actual model, and they additionally provide a check-back whether what we are modeling is consistent and as-intended. Secondly, the example also suggests once more that a game with unawareness is sufficiently described by an exhaustive list of awareness statements and corresponding beliefs. If that is so, $\mathcal{L}^G$, evaluated on suitable awareness structures, has just the right expressivity to mitigate between game model and intuition.

In further defense of this latter claim, let me address two more possible objections. Both concern the idea that perhaps our semantics of $\mathcal{L}^G$ should be made richer to express more fine-grained intuitions about what agents are aware of, and what they believe about others' awareness. Both objections will receive the same line of defense: as far as modeling a game with unawareness is concerned, drawing finer distinctions is not necessary. Moreover, if we stick to the above classical semantics, we can take over standard results from modal model theory without any necessary amendment. This will be crucial in the following, but first here are the possible objections I would like to dismiss:

Firstly, according to our semantics $\mathcal{A}, w \models \Diamond_v(h_1 \vee h_2)$ just in case there is a world $w' \in R_v(w)$ such that either h_1 reoccurs in $L(w')$ or h_2 does. However, intuitively, a sentence of the form "the agent at choice point v is aware of h_1 or h_2" may get a stronger reading, namely that "the agent at choice point v is aware of h_1 *and* she is aware of h_2 in v". This is not a problem exclusively for the present approach but a general problem of providing possible worlds semantics to modals scoping over disjunctions (cf. Kamp 1973). Similar concerns apply to implication under modals in our case. None of this, however, is crucial for our purposes here, because it is not the intention of this construction to give a proper semantics of *all* natural language ascriptions of awareness states. All we need is a proper interpretation for simple awareness statements and beliefs about these. (In fact, we could as well restrict the language $\mathcal{L}^G$ to contain only bare "proposition letters" h, their negation $\neg h$, and conjunctions of these.)

Secondly, the given semantics may appear too simplistic, because it does not acknowledge a certain partiality in belief ascriptions. Obviously, if a belief type $\langle w, v\rangle$ does not occur in $\mathcal{A}_G$, then the above semantics makes all formulas of the form $\Box_v(\varphi)$ true in world w. It therefore seems as if we should rather define a three-valued logic in which the truth value of $\Box_v(\varphi)$ may also be left undefined. However, it is clear that the simpler bivalent logic does not hamper descriptions of games with unawareness in terms of formulas from $\mathcal{L}^G$. For one, we surely do not need the undefinedness of some formula $\Box_v\varphi$ at some world to fully express our intuitions about a game with unawareness. And even if it was necessary to express our intuitions, there would be a way of expressing "undefinedness" also in a bivalent logic, as only for non-occurring belief types will the formula $\Box_v\bot$ be true. (Here, $\bot$ is the "bottom", a formula that is false in every world.) Moreover, it is beneficial to stick to a standard bivalent interpretation, so that we can use standard results of modal model theory without amendement, as we shall see presently.

3.2 Equivalence of Awareness Structures

As we have seen, awareness structures are both representations of games with unawareness *and* models for a simple modal language that *describes* games with unawareness. Relating our game representations to a language therefore helps judging whether the representational format is appropriate. But there is yet another advantage in interpreting awareness structures as models for a modal language. Under the assumption that $\mathcal{L}^G$ is indeed just expressible enough to describe all features of unawareness of moves that we as modellers care about in games with unawareness, the modal model perspective also offers a straightforward, off-the-shelf means of identifying two awareness structures as relevantly equivalent (beyond trivial isomorphism).

Definition 3.5 ((Modal) Equivalence). Fix two awareness structures $\mathcal{A}$ and $\mathcal{A}'$ based on the same game G. We say that $\mathcal{A}$ and $\mathcal{A}'$ are *($\mathcal{L}^G$-)equivalent* $\mathcal{A} \equiv^{\mathcal{L}^G} \mathcal{A}'$ just in case $\mathcal{A}, w_0 \models \varphi$ iff $\mathcal{A}', w_0' \models \varphi$ for all forumlas $\varphi \in \mathcal{L}^G$.

Since $\mathcal{L}^G$ is a standard modal language, and since awareness structures give a standard semantics to expressions in $\mathcal{L}^G$, all of the textbook results about (modal) equivalence apply to our representation of games with unawareness (cf. Blackburn et al. 2001; Blackburn et al. 2007). Modal model theory commands a number of results that relate structural properties of models to modal equivalence in the above sense. To give an impression of what this means, we can construct a *minimal* awareness structure from any *image-finite* awareness structure by a process called *bisimulation contraction*. Here is the idea.

Definition 3.6 (Bisimulation). Fix two awareness structures $\mathcal{A}$ and $\mathcal{A}'$ based on the same game G. A bisimulation between $\mathcal{A}$ and $\mathcal{A}'$ is a relation $Z \subseteq W \times W'$ such that:

(a.) if wZw' then $L(w) = L'(w')$,

(b.) if wZw' and wRu, then there is a $u' \in W'$ such that uZu' and $w'R'u'$, (forth condition)

(c.) if wZw' and $w'R'u'$, then there is a $u \in W$ such that uZu' and wRu . (back condition)

For our purposes, say that $\mathcal{A}$ and $\mathcal{A}$ are bisimilar if there is a bisimulation Z between $\mathcal{A}$ and $\mathcal{A}'$ such that $w_0 Z w_0'$.

An awareness structure is *image-finite* whenever the sets $R_v(w)$ are finite for all w, v. For image-finite $\mathcal{A}$ and $\mathcal{A}'$, it follows from the so-called *Hennessy-Milner theorem* (see Blackburn et al. 2001, page 69) that $\mathcal{A}$ and $\mathcal{A}'$ are bisimilar iff they are modally equivalent.

This gives a straightforward mechanism to compute a minimal representation of a game with unawareness from an image-finite awareness structure $\mathcal{A}$. Take the union of all bisimulations between $\mathcal{A}$ and itself, which is obviously also a bisimulation between $\mathcal{A}$ and itself. This maximal bisimulation on $\mathcal{A}$ identifies equivalence classes of worlds with equal modal properties. *Bisimulation contraction* is the process of forming a new model from these equivalence classes and relating equivalence classes $|w|$ and $|w'|$ with R_v in the reduced structure whenever there are worlds in $|w|$

and $|w'|$ that are related by R_v in the to-be-contracted structure $\mathcal{A}$. The bisimulation contraction of image-finite $\mathcal{A}$ will be the smallest model modally equivalent to $\mathcal{A}$.

If awareness structures are not image-finte, a more complicated structural property ascertains modal equivalence: bisimulation in the models' *ultra-filter extensions*. Sketching this here would lead too far astray, and therefore suffice it to notice that modal model theory may provide us with tools to assess expressibility, equivalence and minimality of *game representations*, which is certainly a very welcome contribution to game theory. (The interested reader is referred to chapter 2 of Blackburn et al. 2001 for details about modal equivalence via ultra-filter extensions.) The notion of equivalence of a game is tied to the set of formulas in $\mathcal{L}^G$ that describes the situation from an intuitive point of view.

3.3 Further Properties of Awareness Structures

Awareness structures conservatively extend classical game models. A classical game G is modeled by any awareness structure for which $L(w) = G$ for all of worlds w. It is easy to check that in any such structure, the game G will be common knowledge in the sense that all awareness statements of the form $\Diamond_v h$ are common knowledge. There are infinitely many such awareness structures, all of which are modally equivalent. But, of course, there is an obvious minimal representation with just one world w_0, $L(w_0) = G$ with $R_v(w_0) = \{w_0\}$ for all v.

Unlike previous functional approaches (Halpern and Rêgo 2006; Feinberg 2009; Heifetz et al. 2010), relational awareness structures can straightforwardly express uncertainty of another agent's awareness, without relegating this to additional uncertainty in the underlying game itself. We have seen an example of this in Figure 2. In this context, consider an additional constraint on awareness structures $\mathcal{A}_G$:

Certainty: for all belief types $\langle w, v\rangle \in B_{\mathcal{A}_G}$, the set $R_v(w)$ is singleton.

This constraint rules out uncertainty of another agent's awareness. The interested reader is encouraged to check that awareness structures with this constraint are effectively equivalent to Feinberg's game models with unawareness Feinberg 2009. (There are only minor differences that relate to the present definition of pruning.)

Any non-trivial awareness structure $\mathcal{A}_G$ embeds a hierarchy of (smaller) awareness structures, in the following sense. For any world w from $\mathcal{A}_G$, we can look at the *w-generated submodel* $\mathcal{A}'$ constructed from $\mathcal{A}_G$ as follows:

- $w_0' = w$,
- $W' = R^*(w)$
- $R_v' = R_v \restriction W'$ for all $v \in L(w)$
- $L' = L \restriction W'$.

In words, we take w as the new actual world, collect all worlds accessible from w in $\mathcal{A}$ and restrict R_v and L accordingly. Although $\mathcal{A}'$ is based on $L(w)$, and this need not be identical to G, it is nonetheless obvious that all reoccurring information states $v \in L(w)$ refer uniquely to information states in $L(w)$ and all its partial representations. Modulo naming of information states, it is easy to check that this construction satisfies all the constraints of Definition 3.1.[6] In effect, we may look at an awareness structure as a hierarchy of games with unawareness, in which subjective views subsume subjective views.

Looking at awareness structures as comprising other awareness structures is helpful when thinking about solutions to games with unawareness. Above we noticed that a classical game G is represented as an awareness structure with $L(w) = G$ for all worlds w. With this, it makes sense to say that an awareness structure $\mathcal{A}_G$ *contains* a classical game $G' \sqsubseteq G$ whenever there is a world w in $\mathcal{A}_G$ with $L(w) = G'$ such that the w-generated submodel is modally equivalent to the classical game G'. If a sequence of view $v_1, \ldots, v_n$ leads to a world w whose w-generated submodel is modally equivalent to a classical game, we say that that sequence *terminates* in a classical game. We can then say that an unawareness structure *terminates* in classical games if there is a finite n such that any sequence of views with length n terminates in a classical game.

Not all awareness structures will have this property or be modally equivalent to one that does. (Feinberg gives an example of a game with unawareness that would afford an infinite, non-terminating succession of worlds and views (Feinberg 2009).) However, plausibly, most practically relevant unawareness structures will have this property, and those that do may lend themselves to a rather naïve algorithmic solution procedure based on any solution for the terminal classical games, as we will see in Section 4.2. But, in order to properly speak properly of solutions for games with unawareness, we should take the step to full-fledged modal game models.

[6]This also follows from the fact that w-generated submodels preserves modal satisfaction in world w (cf. Blackburn et al. 2001, p.56).

4 Modal Game Models

The main objective of this section is to show how awareness structures are easily extended to full epistemic models for reasoning about games with unawareness. These models help study the implications of rationality and various forms of belief in rationality in the class of games with unawareness of moves. This section argues tentatively that characterizations of equilibrium as rational choices under true belief are problematic for games with unawareness. Contrary to that, solutions that select for behavior compatible with common belief in rationality, or certain variations of it, seem reasonable and, replicating similar results of Feinberg 2009 and Heifetz et al. 2010 in terms of modal models, are shown always to exist.

In order to represent the players' behavior and their behavioral beliefs, awareness structures readily extend to full modal game models similar to those proposed by Robert Stalnaker (Stalnaker 1994, 1998). Whether a given modal game model is a viable extension of a given awareness structure is tied to the notion of modal equivalence. We should conceive of awareness structures, then, as something like a scaffolding which defines the possible beliefs of players which a full epistemic model will fill in accordingly.

Definition 4.1 (Modal Game Model). Fix an awareness structure $\mathcal{A}_G$. (For convenience, assume that $\mathcal{A}_G$ is image-finite.) A game model for $\mathcal{A}_G$ is a structure

$$\mathcal{M}_{\mathcal{A}_G} = \langle W, w_0, \{R_v, P_v\}_{v \in V}, L, \{\sigma_w\}_{w \in W} \rangle ,$$

such that $\langle W, w_0, \{R_v\}_{v \in V}, L \rangle$ is an awareness structure that is modally equivalent to $\mathcal{A}_G$, $P_v : \mathcal{P}(W) \to \mathbb{R}^{\geq 0}$ is an additive measure function for each view v that assigns a non-negative level of credence to each proposition, i.e., subset of possible worlds, and σ_w is a behavioral strategy profile for game $L(w)$ such that for all $w' \in R_v(w)$: $\sigma_w(v) = \sigma_{w'}(v)$.[7] We explicitly allow for models with $P_v(\{w\}) = 0$ for some w and v (for reasons that will become clear), but we require that if $R_v(w) \neq \emptyset$ then $P_v(R_v(w)) > 0$.

A model then incorporates or rather extends its underlying awareness structure, but it need not be isomorphic to it; in fact, it will often add worlds for a specification of the players' beliefs about opponent behavior and beliefs; but when it does, then it is constrained so as to preserve all relations of (beliefs about beliefs about ...) awareness. This is then where the connection with the language $\mathcal{L}^G$ comes in once again, and where the notion of modal equivalence is crucial: it is needed to determine when a larger structure, the model, preserves the information of a smaller one, the awareness structure.

Just as for awareness structures, define the set of all *belief types* of a model $\mathcal{M}$ as the set $B_{\mathcal{M}} = \{\langle w, v \rangle \mid R_v(w) \neq \emptyset\}$. A model fixes the behavior of all its belief types $\langle w, v \rangle$ as a probability distribution $\sigma_w(v)$ over successors of v in $L(w)$. This could be interpreted as a strategy even in the game G, not only in $L(w)$. But, most importantly, it must be faithful to the local subjective view of the agent, because we require that $\sigma_w(w) = \sigma_{w'}(v)$ for all $w' \in R_v(w)$. Effectively this means that (i) an agent cannot play actions she is unaware of and that (ii) every agent knows her probabilistic choice. In any model, both of these statements are common knowledge between players (see below for a definition of common knowledge in game models).

A model also fixes the probabilistic beliefs of all belief types in the usual way. For proposition $\varphi \subseteq W$ define the *probabilistic belief* of type $\langle w, v \rangle$ as:

$$\pi_{w,v}(\varphi) = \frac{P_v(R_v(w) \cap \varphi)}{P_v(R_v(w))} .$$

It may be the case that an accessible world $w' \in R_v(w)$ is assigned credence $P_v(\{w'\}) = 0$. Although Stalnaker's models do not allow for this, it is reasonable here, because we would like to ask whether reasoning about rationality could rule out uncertainty about an opponent's awareness. (See example in Figure 5 and the relevant discussion below.)

Therefore define an amended accessibility relation, R_v^+ such that $w' \in R_v^+(w)$ iff $w' \in R_v(w)$ and $P_v(\{w'\}) \neq 0$. With this, say that the belief type $\langle w, v \rangle$ believes a proposition φ in $\mathcal{M}$, iff $R_v^+(w) \subseteq \varphi$. That the player at v believes in proposition φ is itself a proposition, denoted by

$$\mathrm{Bel}_v \varphi = \{w \in W \mid \langle w, v \rangle \in B_{\mathcal{M}} \text{ and } R_v^+(w) \subseteq \varphi\} .$$

[7] To write $\sigma_w(v) = \sigma_{w'}(v)$ is sloppy notation. Certainly, $L(w')$ may contain only a subset of the successor of v that $L(w)$ contains. In that case, the requirement is that $\sigma_w(v)$ places the same probability as $\sigma_{w'}(v)$ on all successors that do also occur in $L(w')$.

That every view point believes in φ is the proposition $\bigcap_{v \in V} \mathrm{Bel}_v \varphi$. Common belief of proposition φ in w is defined as usual as truth everywhere in the transitive closure of R_v^+ starting from w: $\mathrm{CB}\varphi = \{w \in W \mid R_v^{+*}(w) \subseteq \varphi\}$. A proposition φ is true in world w if $w \in \varphi$. As before, knowledge is true belief, and common knowledge is true common belief.

One crucial set of propositions that agents have probabilistic beliefs about is the behavior of all agents (including correct beliefs about their own behavior). In particular, a game model assigns to each of its belief types $\langle w, v \rangle$ a belief in the form of a behavioral strategy profile $\sigma_{\langle w,v \rangle}$ given by:

$$\sigma_{\langle w,v \rangle}(v', a) = \sum_{w' \in W} \pi_{w,v}(\{w'\}) \times \sigma_{w'}(v', a)\,.$$

With this, we say that a belief type $\langle w, v \rangle$ is *rational* in $\mathcal{M}$ iff $\sigma_w(v)$ is a best response in game $L(w, v)$ to the belief $\sigma_{\langle w,v \rangle}$. A player i is rational in world w iff $\langle w, v \rangle$ is rational for all relevant belief types with v a choice of player i. This is then a notion of "subjective rationality": the choice of the agent must be optimal in the game *as she sees it*, not in the game $L(w)$. This is basically where a notion of unawareness in games gets its bite from: even though all players are rational within their subjective representation of a game, their choices need not be objectively rational. The proposition that the player at choice point v is rational is:

$$\mathrm{Rat}_i = \{w \in W \mid \text{player } i \text{ moves at } v \text{ in } L(w) \;\rightarrow \langle w, v \rangle \;\text{ is rational}\}$$

and the proposition that everybody is rational is $\mathrm{Rat} = \bigcap_{i<n} \mathrm{Rat}_i$. The proposition that there is common belief in rationality is CB Rat. The following claims are fairly easy to prove.

Claim 4.2. If the player at v believes that she is rational in w, then $\langle w, v \rangle$ *is* rational.

Claim 4.3. There is common belief of rationality at world w iff all belief types $\langle w', v \rangle \in B_{\mathcal{M}}$ with $w' \in R^{+*}(w)$ are rational.

We are interested in the implications of rationality and various forms of belief in it at the actual world w_0. The question is what does the behavior at w_0 look like in *all* models for a given game with unawareness that satisfy additional constraints on rationality and beliefs.

4.1 Equilibrium as Rational under Mutually True Belief

For a start, consider the simplest epistemic interpretation of equilibrium in a two-player static game as a set strategies both of which are mutually known, and both of which are rational given this knowledge (cf. Stalnaker 1994; Aumann and Brandenburger 1995). The single most striking fact is that, under this characterization, not every game with unawareness has an equilibrium in this sense. The reason is that mutually true beliefs and mutual rationality may be in conflict, for instance, if some agent is not aware of another player's awareness of a dominant choice.

Of course, this depends in part on how we define what it means for a behavioral belief to be true. For our purposes here, the following is a reasonable definition that should be uncontroversial against the background of simultaneous move games. Say that a belief type has a *true* behavioral belief if $\sigma_{\langle w,v \rangle} = \sigma_w$. So, in particular, for a behavioral belief to be true, the belief type $\langle w, v \rangle$ has to be aware of all choice nodes that occur in $L(w)$ and no actions that $\langle w, v \rangle$ is unaware of may be played with positive probability in w.

Here is a non-trivial example that illustrates that mutually true beliefs and rationality can be inconsistent assumptions in games with unawareness. Consider the strategic game G_1 and the pruned version G_2 where column player only has one choice:

G_1	l	r
u	0,2	2,0
d	1,1	1,1

G_2	r
u	2,0
d	1,1

Let i be the row player and j be the column player, and assume that the game with unawareness is given by the following awareness structure $\mathcal{A}_{G_1}$:[8]

$$w_0 : \; G_1 \xrightarrow{\quad i \quad} w_1 : \; G_1 \xrightarrow{\quad j \quad} w_2 : \; G_2$$

[8] Strictly speaking, we have not defined unawareness structures for simultaneous move games, but it is clear how the definition would have to be amended to cover this basic example.

So, in this game with unawareness player i incorrectly believes that j is unaware of action l.

Claim 4.4. There is no model based on awareness structure $\mathcal{A}_{G_1}$ such that at w_0 player j is rational and player i has a true belief about player j's strategy.

Proof. There are infinitely many models based on $\mathcal{A}_{G_1}$ but all of these have to contain an awareness structure modally equivalent to $\mathcal{A}_{G_1}$. We may therefore think of the worlds in any model $\mathcal{M}$ based on $\mathcal{A}_{G_1}$ as falling within (modal) equivalence classes $|w_0|$, $|w_1|$ and $|w_2|$ based on modal equivalence with the respective worlds in $\mathcal{A}_{G_1}$. Obviously, the actual world w_0' in $\mathcal{M}$ will be in $|w_0|$.

Action l strictly dominates r for player j, and she is aware of it in w_0. So, if we assume that in model $\mathcal{M}_{\mathcal{A}_{G_1}}$, player j is rational in w_0', then we have to assign $\sigma_{w_0'}(j, l) = 1$. Additionally, player i's behavioral beliefs in w_0' are given by $\sigma_{\langle w_0', i\rangle}(j)$. This however, is computed based on the strategies assigned to j in worlds in $|w_1|$ and so these will have to conform to the actions available in worlds in $|w_2|$. The only behavioral belief possible for i in w_0' is therefore $\sigma_{\langle w_0', i\rangle}(j, r) = 1$. □

This shows that games with unawareness need not have equilibria, if understood in this way. The problem highlighted by the example is simply that some behavioral beliefs are not possible in games with unawareness. Certainly, we can weaken the epistemic condition for equilibrium in games with unawareness in several ways. But the problem then is to find a reasonable weaker requirement that still preserves the spirit of equilibrium of mutually best responses that would perpetuate under knowledge at w_0. Weaker notions are liable to be too weak in this respect.

Suppose, for instance, that instead of requiring truth in the actual world, we require truthful behavioral beliefs only in those worlds that correspond to classical games (cf. Feinberg 2009). This may seem reasonable, because in those worlds no behavioral beliefs are blocked by unawareness. What kind of behavior would this select for, if we would require rationality in the actual world? The answer is simple: there will always be a solution, but in some games with unawareness this weaker requirement of true behavioral beliefs is entirely vacuous. In some of these cases the solution, i.e., the behavior that is selected for in the actual world of all models, is only constrained by rationality at w_0, and so may be nothing more than a set of strategies none of which is strictly dominated according to the subjective views of players at w_0. To see this, take an arbitrary game G and an awareness structure in which every v conceptualizes the game with just one choice for each $v' \neq v$.

Other variations of the truthfulness requirement may be conceivable, and the jury is still out whether some reasonable characterization of equilibrium along these lines exists for games with unawareness. Based on the considerations above, however, it seems that equilibrium, characterized somehow by rationality and *truthful* behavioral beliefs, does not square well with games of unawareness. The problem certainly is the truthfulness requirement in structures that may exclude certain beliefs.[9] A radical moral that could be drawn from this is that, perhaps, (epistemic interpretations of) solutions for games with unawareness should not require truthfulness at all, but should rather strengthen the impact of rationality, so as to require higher-order belief in rationality.

4.2 Common Belief in Rationality

It follows from Claim 4.3 that if a model $\mathcal{M}$ is connected —and it suffices to restrict attention to these— there is common belief in rationality in actual world w_0 iff *all* belief types of $\mathcal{M}$ are rational. So what does common belief in rationality imply for the behavior of players in w_0? First of all, we should take note that common belief in rationality is never an inconsistent assumption for games with unawareness based on a finite game.

Proposition 4.5. For any awareness structure $\mathcal{A}_G$ based on a finite game G there is always a model $\mathcal{M}_{\mathcal{A}_G}$ with common belief in rationality at w_0 of the model.

The proof, given in the appendix, makes use of another basic fact known from modal model theory, namely the fact that every pointed modal model, and hence every awareness structure too, can be unfolded into a modally equivalent model that is a tree (see Blackburn et al. 2001 p. 62-63). If $\mathcal{A}$ is an awareness structure, then let $T(\mathcal{A})$ be the *unfolded tree* derived from $\mathcal{A}$. The idea is so intuitive that a formal description of the construction is unnecessary: the model $T(\mathcal{A})$ is obtained from $\mathcal{A}$ by treating all paths through $\mathcal{A}$ as worlds of the model $T(\mathcal{A})$ with accessibility relations and assignments $L(\cdot)$ preserved in the obvious way.

[9] None of this says anything yet about the appropriateness of equilibrium for games with unawareness under diachronic interpretations, e.g., as the outcome of learning or population dynamics. Here, a convincing case would have to be made how an agent, or certain relevant parts of a population could remain unaware of certain parts of the game, although the game is played repeatedly. Whether any such interpretation is reasonable will also hinge on the kind of the agents' unawareness, e.g., whether it is more conceptual or rather attentional in nature.

The unfolded tree representation of an awareness structure is helpful also beyond the proof of Proposition 4.5. The set $\mathfrak{M}_{T(\mathcal{A})}$ of all models based on $\mathcal{A}$ that contain $T(\mathcal{A})$ as its awareness structure could be considered something like the class of all *canonical models* for the game $\mathcal{A}$. If we are interested in the implications of rationality and belief in it, then certain possible distinctions in the model will cancel out when we compute a belief types' behavioral beliefs. This makes it clear all distinctions relevant to the (beliefs about) rationality of a choice of a belief type in $\mathcal{A}$ can be expressed by a model in $\mathfrak{M}_{T(\mathcal{A})}$. In other words, when we are interested in the set of solutions selected by, for instance, common belief in rationality, we need to look all models in $\mathfrak{M}_{T(\mathcal{A})}$, but we can safely ignore others.[10]

The canonicity of the tree-based models yields a simple algorithmic way of computing all behavior compatible with common belief in rationality for awareness structures that terminate in classical games by *unravelling*. If $\mathcal{A}$ terminates in classical models, then for each branch in $T(\mathcal{A})$ there is a root-closest world w such that the w-generated submodel is a classical model. The set of strategies compatible with common belief in game $\mathcal{A}$ is obtained from taking all solutions selected by this criterion in the classical games of each branch, and then computing the set of all best responses at each step up to the actual world w_0.

4.3 Belief in Rationality Under Structural Consistency

Rational choice under common belief in rationality is a fairly weak requirement. We can extend the previous modal model constructions to implement additional forward induction reasoning, for instance. This paper does not aspire to go into detail here (but compare the elaborate discussion in Heifetz et al. 2010). We will contend ourselves with a simple construction that allows for a non-trivial application that shows the relation between beliefs expressed in awareness structures, and how these pertain under considerations of rationality. An easy example will show how that uncertainty about awareness in awareness structures may be ruled out by belief in rationality in game models.

In order to extend the idea of extensive form (correlated) rationalizability —with its forward induction rationale (cf. Pearce 1984; Battigalli 1996; Stalnaker 1998)— to epistemic models for games with unawareness in the present setting, we need to strengthen the notion of a possible behavioral belief. Notice simply that, until now, we have allowed a player at choice point v to have behavioral beliefs about opponent behavior that, if correct, would never lead to v in the first place. To strengthen the implications of belief in rationality we would therefore want to place an additional constraint on players' behavioral beliefs in our models, namely that they be *structurally consistent*, at least where this is compatible with higher-order belief in rationality.

Recall that the behavioral belief $\sigma_{\langle w,v\rangle}$ captures what the player at v would believe about her opponents' behavior if world w was actual. Therefore define that $\sigma_{\langle w,v\rangle}$ is *structurally consistent* iff the strategies $\sigma_{w'}$ for all $w' \in R_v(w)$ reach node v with probability 1 in game $L(w)$. The requirement that all belief types have structurally consistent beliefs and that they believe in the opponents' rationality can obviously be inconsistent. Rationality and any higher-order belief in it should be applied only if that is consistent with structural consistency. A version of extensive-form rationalizability for games with unawareness can be defined rather straightforwardly in terms of modal models as follows.

All canonical tree-based models in $\mathfrak{M}_{T(\mathcal{A})}$ not only share the same unawareness structure $T(\mathcal{A})$, but also, since $T(\mathcal{A})$ is a tree, each world $w \neq w_0$ in $T(\mathcal{A})$ identifies a unique belief type in each model. (More than one world may identify a given belief type but that is insubstantial.) It is therefore feasible to associate with each world $w \neq w_0$ from $T(\mathcal{A})$ and model $\mathcal{M} \in \mathfrak{M}_{T(\mathcal{A})}$ the behavioral belief $\mathcal{M}(w)$ that the unique belief type associated with w has in $\mathcal{M}$. So we can say, only somewhat sloppily, that $\mathcal{M}(w)$ is structurally consistent, or that it is a belief in everybody's rationality. Finally, define for a class of models $\mathfrak{M}' \subseteq \mathfrak{M}_{T(\mathcal{A})}$ the set $\mathfrak{M}'(w) = \{\mathcal{M}(w) \mid \mathcal{M} \in \mathfrak{M}'\}$ of behavioral strategies associated with the belief type identified by w in some model in $\mathfrak{M}'$.

Using the canonical tree-based models, we can single out the behavior in games with unawareness that is consistent with any higher-order belief in rationality where this is consistent with structural consistency. Let $\mathfrak{M}^0_{\mathrm{SC}}$ be the set of all tree-based models all of whose types have structurally consistent beliefs. Then define inductively for all $k \geq 1$ (see the proof of Proposition 4.5 for the definition of $R^k(w)$ and $\mathfrak{M}^k$):

$$\mathfrak{M}^k_{\mathrm{SC}} = \left\{\mathcal{M} \in \mathfrak{M}^{k-1}_{\mathrm{SC}} \mid \forall w \in R^k(w_0)\ (\mathfrak{M}^k(w) \cap \mathfrak{M}^k_{\mathrm{SC}}(w) \neq \emptyset \ \rightarrow\ \mathcal{M}(w) \in \mathfrak{M}^k(w))\right\} .$$

The sets $\mathfrak{M}^k_{\mathrm{SC}}$ contain models with beliefs compatible with level-k belief in rationality where this does not violate structural consistency. The behavior in the actual world of all models in the set $\bigcap_k \mathfrak{M}^k_{\mathrm{SC}}$ is the set of behavior that is consistent with common belief in rationality, wherever possible. The proof of the following claim is a simple extension of the proof of proposition 4.5.

Claim 4.6. The set $\bigcap_k \mathfrak{M}^k_{\mathrm{SC}}$ is non-empty for any game with unawareness.

[10] This makes a neat parallel to Feinberg's construction of games with unawareness (Feinberg 2009). Feinberg's infinite collection of finite sequences of views are all the worlds of models in $\mathfrak{M}_{T(\mathcal{A})}$ whenever $\mathcal{A}$ satisfies the certainty constraint.

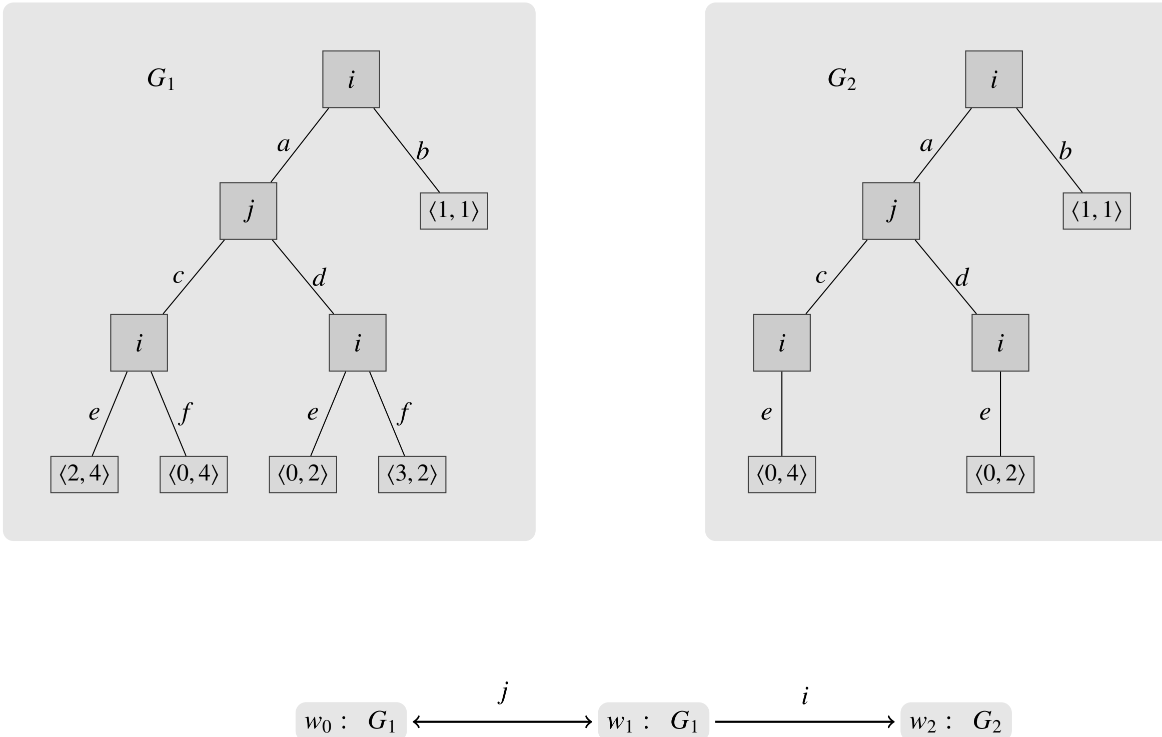

Figure 5: Uncertainty of awareness and forward induction

The following example shows how, although an awareness structure has an agent uncertain about whether another agent is aware of some contingency, the forward induction reasoning captured in this solution can render parts of this uncertainty incompatible with (higher-order) belief in rationality. Consider the game with unawareness in Figure 5. Here, all view points of player i are treated alike. So, the actual game is G_1, but player j is uncertain whether player i is aware of player i's move f at all of her choice points. This uncertainty, although part of the game description, can be leveled based on considerations of rationality. Intuitively speaking, if j has structurally consistent beliefs at her only information state in this game, she must believe that i played a. But then, there is no model in which j can believe that i is rational that assigns positive probability to a world that is modally equivalent to w_2 of the given awareness structure, because in these worlds the choice a is strictly dominated by b. But if j is certain that i is aware of f, her best choice is d, although it is not necessarily when she cannot rule out that i may be unaware. The upshot is that parts of the uncertainty about unawareness of others can be straightened by considerations of rationality in some games. For this to be possible, however, it crucial to allow the definition of models to assign a credence level of zero $P_v(\{w'\}) = 0$ for some worlds $w' \in R_v(w)$.

5 Extensions

The main contribution of this paper is a demonstration of how modal logic can be helpful in the representation of games with unawareness. The last section showed that awareness structures are already halfway towards a modal game model, and that constraints on these models can, if properly executed select reasonable behavioral solutions via assumptions on rationality and belief formation of agents. In the following section I would like to point to a few conceivable extensions of this approach that again speak for a connection of game theory with modal logic in several respects.

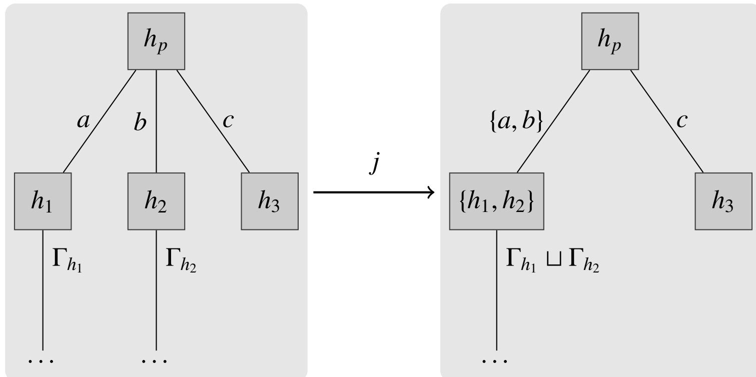

Figure 6: Lumping: Player j cannot distinguish actions a and b

5.1 Unawareness from Lumping

So far, we have only attended to games with unawareness of moves. Consequently, the crucial reduction constraint of our awareness structures regulated that accessible worlds must be partial representations obtained from lumping. We can relax this condition to accommodate other forms of unawareness. Another way in which agents may be unaware in a strategic situation is when they cannot distinguish between several contingencies. This kind of unawareness has so far not been studied systematically in previous models. But it is a very natural one in many situations: some agents may simply not know to make a difference between certain states of nature or action choices.

In order to model unawareness from indistinguishability, we may introduce a second reduction relation obtained by *lumping* branches of the game tree together, while again retaining as much of the labeling as possible. Obviously, not all conceivable ways of such lumping may yield manageable game representations, and these, arguably, do not represent natural cases of unawareness from indistinguishability. For simplicity, we therefore restrict attention to cases where, roughly speaking, what is lumped together are actions of a single player at one of her choice points (plus some more convenient restrictions — see below).

Since a proper definition of lumping is actually rather cumbersome, let us start with a simple intuitive example (see Figure 6). Suppose at history h_p of some game it is player i's turn to choose from successors $H(h_p) = \{h_1, h_2, h_3\}$ labeled by $A(h_p, \cdot)$ as action choices $A(h_p, h_1) = a$, $A(h_p, h_2) = b$, $A(h_p, h_3) = c$. It then may be the case that some player j (not necessarily different from i) can *not* distinguish the actions a and b, but *can* distinguish action c from the former two. (We could think of a European who definitely knows the difference between a handshake and a bow, but cannot distinguish between different types of bowing, that may show different levels of respect in certain Asian cultures.) This can be represented by assigning to j a partial representation of this game in which the histories $H(h_p)$ are lumped together into a set $H'(h_p) = \{\{h_1, h_2\}, h_3\}$ with action labels $A(h_p, \{h_1, h_2\}) = \{a, b\}$, $A(h_p, h_3) = c$. In other words, we will represent those actions that cannot be distinguished as a set of histories, together with the corresponding set of labels as new action label.

To keep things simple, such lumping is only allowed when the play that follows histories h_1 and h_2 in the original game is identical, as far as the player and action labeling is concerned. Conceptually speaking, this is not necessary: it is conceivable, for instance, that an agent could be unable to discriminate actions immediately after which *different* players move. However, excluding this and other aberrant forms of misconceptualization simplifies the definition of lumping.

Definition 5.1 (Lumping). A game Γ' is a *lump* of Γ, $\Gamma' \sqsubseteq_l \Gamma$, iff Γ' is derived from Γ by zero or more successive *lumping steps*. A game Γ' is derived from Γ by one lumping step if Γ' and Γ are identical, except that the branches emanating from two nodes h_1 and h_2 in Γ are replaced with a *merged branch* in Γ'.

Merging branches starting at nodes h_1 and h_2 is only feasible if (i) h_1 and h_2 have an identical predecessor h_p and (ii) the branches are identically labelled except for utilities. More concretely, that means that if $H^{\leq}(h_1) = \{h \in H \mid h_1 \leq h\}$ and $H^{\leq}(h_2) = \{h \in H \mid h_2 \leq h\}$ are the branches emanating from h_1 and h_2 respectively, then we require there to be a bijection $f : H^{\leq}(h_1) \to H^{\leq}(h_2)$ such that for all $h, h' \in H^{\leq}(h_1)$:

- $h < h'$ iff $f(h) < f(h')$
- $P(h) = P(f(h))$

- $A(h, h') = A(f(h), f(h'))$, wherever defined
- if $P(h) = n$, then $\Pr_h = \Pr_{f(h)}$
- if $h_1 < h$, then h and $f(h)$ are in the same information state.

The merged branch $H^{\leq}(h^{\sqcup})$ is then obtained by taking $H^{\leq}(h_1)$ and replacing each node $h \in H^{\leq}(h_1)$ with the set node $\bigcup \{h, f(h)\}$, while retaining the order and labeling in the obvious way. As for the terminal nodes of the merged branch, we assign utilities in lump Γ' that are derived by some "implicit belief". Let $p \in [0; 1]$ and define for all terminal nodes $z^{\sqcup} = \bigcup \{z, f(z)\}$ in the merged branch:

$$u'_i(z^{\sqcup}) = p \times u_i(z) + (1 - p) \times u_i(f(z)) .$$

It remains to be specified how the merged branch is appended to h_p, the predecessor of merged nodes h_1 and h_2 in Γ. The lump Γ' assigns the action labeling

$$A'(h_p, h^{\sqcup}) = \left\{A(h_p, h_1), A(h_p, h_2)\right\}$$

to the maximal element $h^{\sqcup}$ of the merged branch. Moreover, if h_p is a nature move, then we should assign *some* prior probability $\Pr'_{h_p}$ to nature's moves in h_p (see also the discussion below). Finally, we require that if h_1 and h_2 are part of information states v_1 and v_2 respectively, then $h^{\sqcup}$ is part of information state

$$v^{\sqcup} = \{h^{\sqcup}\} \cup v_1 \cup v_2 \setminus \{h_1, h_2\} .$$

This is to ensure that anything (believed) indistinguishable in h_1 and h_2 (is believed to) remain indistinguishable in a representation that lumps h_1 and h_2 together.

The above definition leaves the prior probabilities $\Pr'_{h_p}$ unspecified, when h_p is a nature move. A natural idea would be to conserve the values of $\Pr_{h_p}$ wherever possible, and to assign probability $\Pr'_{h_p}(h^{\sqcup}) = p \times \Pr_{h_p}(h_1) + (1 - p) \times \Pr_{h_p}(h_2)$, where p is again the "implicit belief" hidden in the lumping. This additive assignment of probabilities to sets of states is entirely *deterministic*, in the sense that there is a unique prior that results from lumping (see de Jager 2009 for more on this additive assignment of probabilities of coarse-grained states.) This, however, may occasionally be too inflexible, especially in the light of certain empirical results on subjective probability assignments. Fox and Levav (2004) argue that human laboratory performance in probability assignment tasks, such as, for instance, the Monty-Hall puzzle, can be explained based on the assumption that subjects assign flat probabilities to either naively coarse-grained, or to more sophisticated fine-grained partitions of logical space (cf. Grünwald and Halpern 2003). If we want to allow for flat probabilities irrespective of granularity we should not further restrict the definition of a lumping step across the board.

No matter how probabilities are assigned to indistinguishable states, the lumping relation $\sqsubseteq_l$ also gives rise to a partial order on the set of games $\mathfrak{G}$, just as the pruning relation did. Pruning and lumping are two different, but compatible ways of representing a subjectively restricted conceptualization of a given game. As a single agent may certainly (be believed to) fail in both ways to represent a game correctly, it makes sense to define a partial representation of game G, in the sense of the reduction constraint on awareness structures, as the transitive closure of prunings and lumpings of G.

Other than the meaning of partiality in the reduction constraint, nothing has to change to have awareness structures represent games with unawareness from ignorance and indistinguishability of moves. Notice that it is still possible to identify all information states of any game in any of its pruned-lumped partial representations: even if only some of the nodes in some information state v from game G reoccur partly lumped together in a reduced game G', there will be a unique referent of v in G'.

The only non-trivial amendments in the extension to unawareness from indistinguishability concern (i) an extension of the modal language $\mathcal{L}^G$ to talk about indistinguishability and (ii) the calculation of expected utilities in models, which will have to take into account implicit beliefs hidden in the lumping. It remains to be seen how this form of unawareness behaves under different requirements on rational behavior and beliefs therein. To represent it, awareness structures are certainly flexible and modular enough.

5.2 Awareness of Unawareness

In the present framework, we can easily model an agent's awareness of another agent's unawareness. However, we cannot model an agent's awareness of her *own* unawareness. Halpern & Rêgo have suggested adding *virtual moves* to a game representation that captures the subjective conceptualization of an agent who is aware of her unawareness (cf.

Heifetz et al. 2010). The modal perspective taken in this paper suggests a different and more principled approach. In a recent paper (Halpern and Rêgo 2009), Halpern & Rêgo give an amended propositional modal logic for reasoning about awareness that also allows to reason about one's own unawareness of propositions. Using this, we could construct awareness structures as special models of the extended modal language of Halpern & Rêgo. As far as the game representation is concerned, this approach would be entirely straightforward.

The interesting aspect of this project would be to see how an agent's awareness of her own unawareness behaves under strategic reasoning. The crucial question is what an agent who is aware of her unawareness should do if she observes behavior that is incompatible with the assumed rationality of opponents. Without awareness of unawareness, belief in irrationality of an opponent may be unavoidable. But, depending on the source of unawareness, if unawareness is not conceptual, but rather arising from computational limitations or matters of inattentiveness, then agents should perhaps try to rationalize their observations, if needed, by revision of their conceptual representation of the strategic situation.

The question how, by which process, an agent can actively and consciously become aware of a relevant contingency that could rationalize an observation *by introspection* is beyond the scope of (present-day) rational choice theory. But no matter what the process is, rational choice theory *can* model the revisions necessary for a successful rationalization in terms of awareness dynamics. Here, we could try to exploit the somewhat uncomfortable parallel between belief revision and awareness dynamics (Hill 2010). Modal models of belief revision in terms of orderings of possible worlds abound. More crucially, models where, as would be necessary for revising unawareness, the revisions are not introspectively accessible to an agent at the outset, are also readily available (cf. Board 2004). This, however, is as of now mere speculation and has to be left for another occasion.

6 Conclusion

In conclusion, this paper has suggested to represent games with unawareness in terms of, basically, pointed and relevantly constrained models for a basic modal logic of belief. The notion of modal equivalence tied to the language with which to capture agents' beliefs about unawareness has been crucial in several respects, foremost to extend awareness structures to epistemic models of the game with unawareness. Basic results from modal model theory proved helpful in assessing the implications of rationality and belief in rationality in games with unawareness.

A Proofs

Proof of Proposition 3.2. In Section 2.2 we observed that a pruning of a given game G corresponds one-to-one with a suitable substet $H' \subseteq H$ of nodes. Suitable here means that H' may not introduce new terminal nodes. It is obvious that for all w the set $\{h \in H \mid M, w \models h \rightarrow z(h)\}$ depicts exactly this property.

As for the equivalence of constraints, centering is trivial, and so is reduction. Existence is easily checked when we notice that (i) v^* is true in v iff v reoccurs in $L(w)$ and that (ii) $\Diamond_v \top$ true in a world w iff there $R_v(w) \neq \emptyset$. These two observations also explain the characterization of the relevance constraint. Finally, introspection needs no argument, as this is well known. □

Proof of Proposition 4.5. Fix an arbitrary awareness structure $\mathcal{A}$ and construct from it its unique modally equivalent *unfolded tree* $T(\mathcal{A})$. Let $\mathfrak{M}$ be the set of all models which contain exactly the awareness structure $T(\mathcal{A})$. Since then all these models share the set of possible worlds and accessibility relations, we can safely refer to these as simply W and R_v in the following.

For convenience define different relevant sets of accessible worlds in models from $\mathfrak{M}$ for all $k \geq 1$:

$$R^k(w) = \left\{w' \in W \mid \exists v_1, \ldots, v_k \ : \ w' \in R_{v_1 \ldots v_k}(w)\right\} .$$

The set $R^k(w_0)$ are then the worlds accessible via k steps from the actual world. With this, define inductively the sets $\mathfrak{M}^k$ as the sets of all models from $\mathfrak{M}$ in which there is, what we may call, level-k belief in rationality:

$$\mathfrak{M}^1 = \left\{\mathcal{M} \in \mathfrak{M} \mid R^1(w_0) \subseteq \text{Rat is true in } \mathcal{M}\right\}$$
$$\mathfrak{M}^{k+1} = \left\{\mathcal{M} \in \mathfrak{M}^{k+1} \mid R^k(w_0) \subseteq \text{Rat is true in } \mathcal{M}\right\} .$$

The sets $\mathfrak{M}^k$ are non-empty for any k. To see this, first notice that for any classical game and choice point in that game, there is always a belief and strategy in that choice point such that the strategy is rational given that belief. Now, consider the case $\mathfrak{M}^1$. From the remark we just made, it is clear that, since all models in $\mathfrak{M}$ are based on a tree model $T(\mathcal{A})$, it is always possible to assign strategy profiles to all worlds $w'' \in R^2(w_0)$ such that the behavior of choice point v such that $w' R_v w''$ for the *unique* —as $T(\mathcal{A})$ is a tree— predecessor of w'' is rational in $L(w'')$. The argument for the inductive step is parallel. If $\mathfrak{M}^k$ is non-empty, then so is $\mathfrak{M}^{k+1}$. Take any arbitrary $w \in R^{k+1}(w_0)$ and repeat the argument from the induction base with w instead of w_0.

The remaining question is whether $\bigcap_k \mathfrak{M}^k$ is non-empty. Define the sets S^k as the set of all pure strategies that lie in the support of w_0 for some model in $\mathfrak{M}^k$. Obviously, $S^{k+1} \subseteq S^k$, but also $S^k \neq \emptyset$ for all k. Remains to notice that the set of probabilistic strategies that are rational under some belief are completely determined by the pure strategies that are. Since there are only finitely many pure strategies in a finte game G the result follows. □

Acknowledgements

I would like to thank Yossi Feinberg, Gerhard Jäger, Sven Lauer, Martin Meier, Roland Mühlenbernd and Jason Quinley, for discussions and stimulating comments. Also, I would like to especially express my appreciation for many inspiring hours of discussing unawareness with Tikitu de Jager. All errors are my own.

References

Aumann, Robert and Adam Brandenburger (1995). “Epistemic Conditions for Nash Equilibrium”. In: *Econometrica* 63.5. Pp. 1161–1180.

Battigalli, Pierpaolo (1996). “Strategic Rationality Orderings and the Best Rationalization Principle”. In: *Games and Economic Behavior* 13. Pp. 178–200.

Blackburn, Patrick et al. (2001). *Modal Logic*. Cambridge Universtiy Press.

Blackburn, Patrick et al., eds. (2007). *Handbook of Modal Logic*. Studies in Logic and Practical Reasoning. Elsevier.

Board, Oliver (2004). “Dynamic Interactive Epistemology”. In: *Games and Economic Behavior* 49. Pp. 49–80.

Dekel, Eddie et al. (1998). “Standard State-Space Models Preclude Unawareness”. In: *Econometrica* 66.1. Pp. 159–173.

Fagin, Ronald and Joseph Y. Halpern (1988). “Belief, Awareness and Limited Reasoning”. In: *Artificial Intelligence* 34. Pp. 39–76.

Feinberg, Yossi (2004). “Subjective Reasoning — Games with Unawareness”. Research Paper No. 1875, Stanford University.

— (2009). “Games with Unawareness”. Unpublished manuscript, Stanford University.

Fox, Craig R. and Jonathan Levav (2004). “Partition-Edit-Count: Naive Extensional Reasoning in Judgement of Conditional Probability”. In: *Journal of Experimental Psychology: General* 133.4. Pp. 626–642.

Grünwald, Peter and Joseph Y. Halpern (2003). “Updating Probabilities”. In: *Journal of Artificial Intelligence Research* 19. Pp. 243–278.

Halpern, Joseph Y. (2001). “Alternative Semantics for Unawareness”. In: *Games and Economic Behavior* 37. Pp. 321–339.

— (2003). *Reasoning about Uncertainty*. MIT Press.

Halpern, Joseph Y. and Leandro Chaves Rêgo (2006). “Extensive Games with Possibly Unaware Players”. In: *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*. Pp. 744–751.

— (2008). “Interactive Unawareness Revisited”. In: *Games and Economic Behavior* 62. Pp. 232–262.

— (2009). “Reasoning about Knowledge of Unawareness”. In: *Games and Economic Behavior* 67. Pp. 503–525.

Heifetz, Aviad et al. (2006). “Interactive Unawareness”. In: *Journal of Economic Theory* 130. Pp. 78–94.

— (2008). “A Canonical Model for Interactive Unawareness”. In: *Games and Economic Behavior* 62. Pp. 304–324.

Heifetz, Aviad et al. (2010). “Dynamic Unawareness and Rationalizable Behavior”. Unpublished manuscript.

Hill, Brian (2010). “Awareness Dynamics”. In: *Journal of Philosophical Logic* 39. Pp. 113–137.

de Jager, Tikitu (2009). “ ‘Now that you mention it I wonder. . . ’: Awareness, Attention, Assumption”. PhD thesis. Universiteit van Amsterdam.

Kamp, Hans (1973). “Free Choice Permission”. In: *Proceedings of the Aristotelian Society* 74. Pp. 57–74.

Kreps, David M. and Robert Wilson (1982). “Sequential Equilibria”. In: *Econometrica* 50.4. Pp. 863–894.

Li, Jing (2009). “Information Structures with Unawareness”. In: *Journal of Economic Theory* 144.3. Pp. 977–993.

Modica, Salvatore and Aldo Rustichini (1999). “Unawareness and Partitional Information Structures”. In: *Games and Economic Behavior* 27. Pp. 265–298.

Ozbay, Erkut Y. (2007). “Unawareness and Strategic Announcements in Games with Uncertainty”. In: *Proceedings of TARK XI*. Ed. by Dov Samet. Pp. 231–238.

Pearce, David G. (1984). “Rationalizable Strategic Behavior and the Problem of Perfection”. In: *Econometrica* 52.4. Pp. 1029–1050.

Stalnaker, Robert (1994). “On the Evaluation of Solution Concepts”. In: *Theory and Decision* 37. Pp. 49–73.

— (1998). “Belief Revision in Games: Forward and Backward Induction”. In: *Mathematical Social Sciences* 36. Pp. 31–56.