Dynamic Awareness

Joesph Y. Halpern[1] , **Evan Piermont**[2]
[1]Cornell University Computer Science Department
[2]Royal Holloway, University of London, Department of Economics
halpern@cs.cornell.edu, evan.piermont@rhul.ac.uk

Abstract

We investigate how to model the beliefs of an agent who becomes more aware. We use the framework of Halpern and Rêgo (2013) by adding probability, and define a notion of a *model transition* that describes constraints on how, if an agent becomes aware of a new formula φ in state s of a model M, she transitions to state s^* in a model M^*. We then discuss how such a model can be applied to *information disclosure*.

1 Introduction

Agents must sometimes take actions in situations that they do not fully comprehend. Work in computer science, economics, and philosophy has sought to capture this formally by considering agents who are *unaware* of some aspects of the world. (See, e.g., (Fagin and Halpern 1988; Halpern and Rêgo 2013; Modica and Rustichini 1994; 1999; Heifetz *et al.* 2006; Board and Chung 2009; Sillari 2008).) Most work on awareness thus far has focused on the static case, where awareness does not change. The focus of this paper is the dynamic case: how to how to model the beliefs of an agent who becomes *more* aware.

When there is no unawareness, the standard approach to dealing with beliefs is well understood; we update by conditioning. However, it is far from clear how beliefs should change when an agent becomes aware of new features. For example, a mathematician who becomes aware of the Riemann hypothesis, but learns nothing beyond that, may change his methods of proof and his beliefs about what is true, despite not having learned that any particular event has obtained. When an agent becomes more aware, his entire (subjective) view may change. In this paper, we propose a model for the dynamics of increased awareness for agents who are introspective, and so can reason about their own and other's awareness (and lack of it).

We work with a probabilistic extension of the framework of Halpern and Rêgo (2013) where agents understand that they themselves may be unaware of some propositions, but cannot directly reason about or articulate such propositions. Instead the agents consider states (possible worlds) that contain *shadow propositions*—proxies that represent an agent's vague conception of those propositions of which he is unaware. We define a notion of a *model transition* that describes constraints on how, if an agent becomes aware of a new formula φ in state s of a model M, she transitions to state s^* in a model M^*. When an agent becomes aware of a formula that she was previously unaware of, she associates the novel propositions with shadow propositions she previously considered. The states of the updated model resemble the original states except that some shadow propositions may be replaced by the novel propositions contained in the newly discovered formula.

The transition rule allows uncertainty to both increase (because the agent can be uncertain about how to interpret the new propositions) and decrease (because by making a discovery, the agent learns about his own unawareness). This is in contrast to *reverse Bayesianism* which, roughly speaking, requires that if the agent is aware of both ψ and ψ', and becomes aware of φ, then his assessment regarding the relative likelihood of ψ and ψ' should not change (Karni and Vierø 2013). Karni and Vierø's requirement does not seem appropriate for many situations of interest, especially for introspective agents, who may believe that the mere existence of φ is itself informative about the world—so becoming aware of φ changes beliefs about other propositions.

For example, suppose that agent i believes it is very likely that he is fully aware and observes j taking questionable actions. He might then reasonably believe that agent j is not rational. But then suppose that i becomes aware of some fact φ of which he was not previously aware. Since he now acknowledges that he did not fully understand the situation, i considers it more likely that j can rationalize his behavior. Hence, simply by becoming aware of φ, i changes his assessment of j being rational.

We then discuss how such a model can be applied to *information disclosure*. Many economic models predict voluntary disclosure due to strategic concerns (because no news is interpreted as bad news). For example, consider restaurant health code inspections, which rate restaurants on a scale of A-F. Surely, an A restaurant would display the rating in the window, so a rational and fully aware patron who sees a restaurant without a rating can assume that the restaurant does not have

an A rating. But then, B restaurants would want their rating to be known, rather than having patrons wonder where in the B to F range they fall; hence B restaurants will also post their ratings. Continuing inductively, it follows that all restaurants will disclose their ratings (except possible those with an F). Economists have used such analyses to conclude that many types of financial disclosures will be made voluntarily, so regulation is not needed.

However, when agents might be unaware of the rating system, this analysis clearly fails. An agent unaware of the rating system can draw no conclusion whatsoever from the lack of a rating. Since disclosure might not only provide information but also expand awareness, belief and behavior might change in subtle ways. For example, we show that whether a restaurateur discloses a rating is determined by his beliefs about how patrons will interpret the rating if they were unaware of the system. In particular, if the restaurateur believes that patrons view a posted grade as arising from a binary pass-fail system, then the B restaurant might withhold its rating, despite B objectively being an above-average rating.

More generally, there is a growing experimental literature suggesting that subjects often fail to properly reason about hypothetical or counterfactual events, and thus deviate from the predictions of rational (and omniscient) agents; see, for example, (Esponda and Vespa 2014; John *et al.* 2016; Cox *et al.* 2017; Jin *et al.* 2018). Such deviations need not result from wholesale irrationality, but could arise from subjects' unawareness of the relevant counterfactuals, for example, being unaware of the actions an opponent could have taken but did not. Our model provides a framework for considering unawareness in a strategic setting, and in particular, about the effect of agents' beliefs about other agents' reactions to becoming more aware.

The rest of this paper is organized as follows: Section 2 introduces the syntax and semantics of our static propositional logic. We extend this model to a dynamic environment in Section 3. The key definition is of a *transition rule*, which captures how a model can change when an agent becomes aware of a new formula. Section 4 shows how thinking in terms of transition rules can be helpful in making predictions in economic environments of information disclosure, where one agent can strategically decide to make another agent aware of a new formula. Finally, Section 5 concludes.

2 A (Static) Propositional Logic of Awareness and Probabilistic Beliefs

In this section, we introduce a propositional logic of awareness and probability that essentially combines ideas from the propositional logic of awareness introduced by Halpern and Rêgo (2013) (henceforth HR13) and the logic of knowledge and probability introduced by Fagin and Halpern (1994) (henceforth FH94). The syntax of the logic is as follows: given a set $\{1, \ldots, n\}$ of agents, formulas are formed by starting with a countably infinite set $\Phi = \{p_1, p_2, \ldots\}$ of primitive propositions and a countably infinite set $\mathscr{X}$ of variables, and then closing off under conjunction ($\wedge$), negation ($\neg$), the modal operators A_i, X_i, $i = 1, \ldots, n$, *likelihood formulas* of the form $a_1\ell_i(\varphi_1) + \cdots + a_k\ell_i(\varphi_k) > b$, where $a_1, \ldots, a_k, b$ are rational numbers, and quantification, so that if φ is a formula, then so is $\forall x\varphi$.

Some comments on the syntax: Following Fagin and Halpern (1988), X_i denotes explicit knowledge (or belief); $X_i\varphi$ is true if, in addition to φ being true in all states that i considers possible, i is aware of φ. We capture awareness using the A_i modality; $A_i\varphi$ means that i is aware of φ. Awareness is syntactic, so i may be aware of p_1 but not aware of $p_1 \wedge (p_2 \vee \neg p_2)$. In this paper, we assume for simplicity that awareness is *generated by primitive propositions* (an assumption that goes back to Fagin and Halpern (1988)), so that an agent is aware of a formula iff he is aware of all the primitive propositions in the formula). Following HR13, we use quantification to capture knowledge of unawareness. For example, the formula $X_i(\exists x \neg A_i(x))$ says that agent i (explicitly) knows that there is a formula that he is unaware of. (As usual, $\exists x\varphi$ is an abbreviation for $\neg\forall x\neg\varphi$.) The ℓ_i in a likelihood formula can be interpreted as probability, so a formula such as $a_1\ell_i(\varphi_1) + a_2\ell_i(\varphi_2) > b$ says that a_1 times the probability of φ_1 plus a_2 times the probability of φ_2 is at least b. Call this language $\mathcal{L}_n^{\forall,X,\ell,A}(\Phi)$.

As in first-order logic, we can define inductively what it means for a variable x to be *free* in a formula φ. Intuitively, an occurrence of a variable is free in a formula if it is not bound by a quantifier. A formula that contains no free variables is called a *sentence*. If ψ is a formula, let $\varphi[x/\psi]$ denote the formula that results by replacing all free occurrences of the variable x in φ by ψ. (If there is no free occurrence of x in φ, then $\varphi[x/\psi] = \varphi$.) Unlike standard quantified modal logic, where the quantifiers range over propositions (intuitively, sets of states), following HR13, here the quantifiers range over quantifier-free sentences. Thus, $\forall x\varphi$ is true iff $\varphi[x/\psi]$ is true for all quantifier-free sentences ψ. Roughly speaking, we want quantification to range over formulas, since A_i is syntactic. However, it cannot range over all formulas, since, for example, the formula $\forall x(x)$ would then be true iff all formulas (including itself) were true, and we would not be able to get a recursive definition of truth. We avoid these difficulties by taking the domain of quantification to be the quantifier-free sentences. (See HR13 for further discussion of these syntactic constraints.)

We give semantics to these formulas in probabilistic awareness structures. Given Φ, let Φ^+ consist of Φ together with an infinite set $\Phi' = \{q_1, q_2, \ldots\}$ (disjoint from Φ) of special primitive propositions that we call *shadow propositions*; we explain their role below. A *probabilistic awareness structure for n agents (over Φ^+)* is a tuple $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \mathscr{PR}_1, \ldots, \mathscr{PR}_n, \mathcal{L})$ satisfying the following properties:

- S is a set of states (or possible worlds), which we take

to be finite for simplicity.

- $\pi : S \times \Phi^+ \rightarrow \{\mathbf{true}, \mathbf{false}\}$ determines which primitive propositions in Φ^+ are true at each state in S.
- $\mathcal{K}_i$ is a binary relation on S for each agent $i = 1, \ldots, n$, which for this paper we take to be *Euclidean* (if $(s_1, s_2) \in \mathcal{K}_i$ and $(s_1, s_3) \in \mathcal{K}_i$, then $(s_2, s_3) \in \mathcal{K}_i$), transitive, and *serial* (for all states s, there exists a state s' such that $(s, s') \in \mathcal{K}_i$).
- $\mathscr{PR}_i$ associates with each state s a probability on the states in $\mathcal{K}_i(s) = \{s' : (s, s') \in \mathcal{K}_i\}$, where all subsets of $\mathcal{K}_i(s)$ are taken to be measurable and $\mathscr{PR}_i(s') = \mathscr{PR}_i(s)$ if $s' \in \mathcal{K}_i(s)$.
- $\mathcal{A}_i$ is a function associating a set of propositions with each state in S, for $i = 1, ..., n$ such that if $s' \in \mathcal{K}_i(s)$, then $\mathcal{A}_i(s) = \mathcal{A}_i(s')$.
- $\mathcal{L}$ associates with each state s a subset of Φ^+; we require that $(\cup_{i=1}^n \mathcal{A}_i(s)) \subseteq \mathcal{L}(s)$ and that if $s' \in \mathcal{K}_i(s)$, then $\mathcal{L}(s') \subseteq \mathcal{A}_i(s) \cup \Phi'$.

Intuitively, if $(s, s') \in \mathcal{K}_i$, then agent i considers state s' possible at s. The assumption that $\mathcal{K}_i$ is Euclidean, transitive, and serial means that the K_i operator satisfies the axioms of KD45 traditionally associated with the logic of belief (see (Fagin *et al.* 1995) for more discussion).[1] $\mathcal{A}_i(s)$ is the set of primitive propositions that agent i is aware of at state s. It is more standard to take $\mathcal{A}_I(s)$ to be the set of *all* formulas that i is aware of at world s. We are assuming here that awareness is generated by primitive propositions (Fagin and Halpern 1988); agent i is aware of a formula φ if i is aware of all the primitive propositions in φ. Thus, it suffices to take $\mathcal{A}_i(s)$ to consist only of primitive propositions. The fact that $\mathcal{A}_i(s') = \mathcal{A}(s)$ if $s' \in \mathcal{K}_i(s)$ means that an agent knows what he is aware of; similarly, the assumption that $\mathscr{PR}(s') = \mathscr{PR}(s)$ if $s' \in \mathcal{K}_i(s)$ means that an agent knows his beliefs.

Intuitively, $\mathcal{L}(s)$ is the language associated with state s. We certainly want the language to include all the primitive propositions that some agent is aware of at state s, but possibly others as well. For example, if agent i knows at state s that there is a formula that no agent is aware of (which can be expressed as $X_i(\exists x(\neg A_1(x) \wedge \ldots \wedge \neg A_n(x)))$, then at each state s' that he considers possible, the language must include a primitive proposition not in $\cup_{j=1}^n \mathcal{A}_j(s)$. Moreover, as noted by HR13, we must allow different languages at different worlds to deal with the possibility that agent i might be unsure of whether he is aware of all formulas (which can be expressed as $\neg X_i \neg(\exists x \neg A_i(x)) \wedge \neg X_i \neg(\forall x(A_i(x))))$. In this case, there must be states s' and s'' in $\mathcal{K}_i(s)$ such that $\mathcal{L}(s') = \mathcal{A}_i(s)$ and $\mathcal{L}(s'') \supset \mathcal{A}_i(s)$ (where $\supset$ denotes strict superset). If $s' \in \mathcal{K}_i(s)$, then we think of the primitive propositions in $\mathcal{L}(s') - \mathcal{A}_i(s)$ as "shadow propositions". Agent i understands that they must exist, at some level, but since i is not aware of them, he does not really understand what they denote. Thus, we require that $\mathcal{L}(s')$ consist of the "real" primitive propositions that i is aware of and possibly some shadow propositions. HR13 did not distinguish real and shadow propositions. Making the distinction has no effect on their axioms involving awareness, but we find it conceptually useful, especially to discuss belief dynamics in the next section. However, it is worth noting that this requirement means that, in general, $\mathcal{K}_i$ will not be reflexive. If $\mathcal{L}(s)$ contains real primitive propositions that agent i is unaware of, then we cannot have $s \in \mathcal{K}_i(s)$.

We now define what it means for a formula to be true at a state s in a probabilistic awareness structure M by combining the earlier definitions of FH94 and HR13. We take $\mathcal{L}_n^{\forall,X,\ell,A}(s)$ to consist of the formulas all of whose primitive propositions are in $\mathcal{L}(s)$. For a formula φ let PROP(φ) denote the set of primitive propositions in the formula φ.

- $(M, s) \models p$ if $p \in \mathcal{L}(s)$ and $\pi(s, p) = \mathbf{true}$;
- $(M, s) \models \neg\varphi$ if $\varphi \in \mathcal{L}_n^{\forall,X,\ell,A}(s))$ and $(M, s) \not\models \varphi$;
- $(M, s) \models \varphi \wedge \psi$ if $(M, s) \models \varphi$ and $(M, s) \models \psi$;
- $(M, s) \models A_i\varphi$ if PROP(φ) $\in \mathcal{A}_i(s)$;
- $(M, s) \models X_i\varphi$ if $(M, s) \models A_i\varphi$ and $(M, s') \models \varphi$ for all $s' \in \mathcal{K}_i(s)$;
- $(M, s) \models \forall x\varphi$ if $(M, s) \models \varphi[x/\psi]$ for all quantifier-free sentences $\psi \in \mathcal{L}_n^{\forall,X,\ell,A}(s)$;
- $(M, s) \models a_1\ell_i(\varphi_1) + \cdots + a_k\ell_i(\varphi_k) > b$ if $a_1\mathscr{PR}_i(s)([\![\varphi_1]\!]_M \cap \mathcal{K}_i(s)) + \cdots + a_k\mathscr{PR}_i(s)([\![\varphi_k]\!]_M \cap \mathcal{K}_i(s)) > b$, where $[\![\varphi]\!]_M = \{s' : (M, s') \models \varphi\}$, and $(M, s) \models A_i(\varphi_1 \wedge \ldots \wedge \varphi_k)$.

Example 1. *Consider a model with two agents, i and j, and three states, s_1, s_2, and s_3. $\mathcal{L}(s_1)$ consists of two propositions, p and p', both "real". In state s_1, agent i is unaware of p', while j is aware of both. $\mathcal{L}(s_2) = \{p, q\}$, where q is a shadow proposition. In state s_2, agent j is aware of both p and q, while i is aware only of p. Finally, $\mathcal{L}(s_3) = \{p\}$, and both agents are aware of p in state s_3. $\mathcal{K}_i(s_k) = \{s_2, s_3\}$ and $K_j(s_k) = s_k$ for $k = 1, 2, 3$.*

In state s_1, i believes its possible that j is aware of something he himself is not aware of. However, he can only describe this state vaguely, envisaging not p', but the shadow proposition q. He also considers it possible that he is fully aware. If p is true at state s_2, then i knows that if he is unaware of something, then p is true. ∎

3 Dynamics

We now turn our attention to how an agent's beliefs should be updated when the agent becomes aware of a formula φ of which he was previously unaware.[2] As

[1] As is often done, we blur the distinction between knowledge and belief in this discussion.

[2] Of course, an agent may become aware of φ by learning φ (i.e., by learning that φ is true). We view this conceptually as the composition of two updates: the update due to

the general case is notationally complex, to ease exposition, we first consider the single-agent case. With only a single agent, the main concern is how the agent interprets the formula he becomes aware of, and how knowledge and probabilistic reasoning depend on his interpretation. In the general case, considered in the next subsection, we also need to deal with the beliefs of other agents, and with higher-order beliefs.

3.1 Dynamics: The Single-Agent Case

Let $\mathcal{L}_i^{\forall,X,\ell,A}(\Phi)$ denote the language with a single agent i. Suppose that in state s in a model $M = (S, \pi, \mathcal{K}_i, \mathcal{A}_i, \mathscr{P}\mathscr{R}_i, \mathcal{L})$, the (single) agent i becomes aware of a formula φ of which he was previously unaware. Our goal is to construct an updated model $M^* = (S^*, \pi^*, \mathcal{K}_i^*, \mathcal{A}_i^*, \mathscr{P}\mathscr{R}_i^*, \mathcal{L}^*)$ and state $s^* \in S^*$ that reflects this. Each state in M^* corresponds to some state in M; the correspondence is captured by a relation T. We now explain how T works.

If t is a state such that $\mathcal{L}(t)$ has at least as many shadow propositions that i is unaware of as there are propositions in φ that i is unaware of, then t is *consistent with i becoming aware of φ*. Let $\text{CONS}(M, \varphi, i) \subseteq S$ denote the states consistent with i becoming aware of φ in model M. If $\mathscr{P}\mathscr{R}_i(s)(\text{CONS}(M, \varphi, i)) = 0$, then becoming aware of φ was an event to which i previously gave probability 0; i didn't consider it possible that there were two primitive propositions of which he was unaware. In this case, we place no constraints on how i updates his beliefs; this is analogous to conditioning on an event of measure 0. (We make precise below what "no constraints" means.)

If, on the other hand, $\mathscr{P}\mathscr{R}_i(s)(\text{CONS}(M, \varphi, i)) \neq 0$, then for each $t \in \mathcal{K}_i(s) \cap \text{CONS}(M, \varphi, i)$, i must decide how to take φ into account at t. For example, suppose that at a state s where i is unaware of p and p', i becomes aware of the formula $p \wedge p'$. Further suppose that at $t \in \mathcal{K}_i(s)$, there are exactly three shadow propositions, q, q', and q'', in $\mathcal{L}(t)$. Then i must decide which of q, q', and q'' is p and which is p'; i could in principle decide that none of them is an appropriate candidate for p (or p'). For example, after becoming aware of $p \wedge p'$, i might consider possible a state t^* such that (1) $\mathcal{L}^*(t^*) = \mathcal{L}(t) - \{q, q'\} \cup \{p, p'\}$; (2) $\mathcal{A}_i^*(t^*) = \mathcal{A}_i(s) \cup \{p, p'\}$ (3) $\pi^*(t^*, r) = \pi(t, r)$ if $r \in \mathcal{L}(t) - \{q, q'\}$, $\pi^*(t^*, p) = \pi(t, q)$, and $\pi^*(t^*, p') = \pi(t, q')$. Thus, in t^*, i has replaced the shadow proposition $q \in \mathcal{L}(t)$ with p, and q' with p'. Agent i might also consider possible a state where q' is replaced by p and q'' is replaced by p'. Roughly speaking, in moving from M to M^*, we want to replace each state $t \in \mathcal{K}_i(s)$ by some states compatible with t (each of these states will be related to t by T) and distribute the probability of t among these states, and then condition on $\text{CONS}(M, \varphi, i)$.

becoming aware of φ, followed by the update due to learning φ given that the agent is aware of φ, which can be handled by standard techniques, specifically, conditioning.

The next definition makes precise the relationship between related states.

Definition 1. *$f : \Phi^+ \to \Phi^+$ is a φ-*replacement scheme *if f is the identity on $\Phi^+ - \text{PROP}(\varphi)$ (where $\text{PROP}(\varphi)$ is the set of propositions in φ) and f is injective on $\text{PROP}(\varphi)$. A φ-replacement scheme f is* compatible with i becoming aware of φ at s *if f is the identity on $\text{PROP}(\varphi) \cap \mathcal{A}_i(s)$ and $f(\text{PROP}(\varphi) - \mathcal{A}_i(s)) \subseteq (L(s) \cap \Phi') - \mathcal{A}_i(s)$. A state s^* is an f-*replacement for s *if (1) $f(\mathcal{L}^*(s^*)) = \mathcal{L}(s)$ and (2) $\pi^*(s^*, p) = \pi(s, f(p))$ for all $p \in \mathcal{L}(s^*)$.*

Intuitively, a φ-replacement scheme for i describes how i interprets the propositions in φ of which he was previously unaware, associating each with a different (shadow) proposition. That is, f maps each proposition in φ that i becomes aware of to some (unique) shadow proposition that i was unaware of, and leaves all other propositions alone.

If $f(p) = q$, then i considers it possible that the shadow proposition q represents the real proposition p that appears in φ. So as discussed above, if after becoming aware of $p \wedge p'$, i considers t^* where p was represented by $q \in \mathcal{L}(t)$ and p' by q', then t^* is a f-replacement of t where $f(p) = q$, $f(p') = q'$, and f is the identity otherwise. If $\mathscr{P}\mathscr{R}_i(s)(\text{CONS}(M, \varphi, i)) \neq 0$, then the relation T mentioned above actually associates each state $t^* \in \mathcal{K}_i^*(s^*)$ with a pair (t, f), where $t \in \mathcal{K}_i(s)$ and t^* is an f-replacement of t compatible with i becmoing aware of φ at s.

A *situation* is a pair (M, s) consisting of a model M and a state s in M. Fix a set Φ of primitive propositions and a set $\mathcal{I}$ of agents.[3] Let $\mathcal{M}(\mathcal{I}, \Phi)$ and $\mathscr{S}(\mathcal{I}, \Phi)$ consist of all models and situations, respectively, over the language Φ with agents in $\mathcal{I}$. Let $RS(\varphi, i, s)$ denote the set of φ-replacement schemes compatible with i becoming aware of φ at s, and let $RS(\varphi) = \{id\} \cup (\cup_{i,s} R(\varphi, i, w))$ (where id is the identity function on Φ^+).

The next definition is the one that we have been heading for. It describes what counts as an acceptable update from a situation (M, s) to a new situation (M^*, s^*) that is the result of agent i becoming aware of φ. We capture this using the notion of an acceptable transition rule τ. Formally, a *transition rule* τ maps a situation, a formula, and an agent (for now, necessarily the single agent i) to a new situation, interpreted as the result of agent i becoming aware of the formula φ when the initial situation was (M, s). (M^*, s^*) is an acceptable update from a situation (M, s) after agent i becomes aware of φ if $\tau((M, s), i, \varphi) = (M^*, s^*)$ for some acceptable transition rule τ. Although, conceptually, the ideas behind the definition are quite straightforward, writing them down carefully results in a complicated definition. We give some intuition for the details of the definition

[3] Although this section deals only with the single-agent case, so $\mathcal{I} = \{i\}$, we introduce the more general notation for so that we can use it in later sections.

immediately after the definition, and then provide an example that illustrates some of them.

Definition 2. $\tau : \mathscr{S}(\{i\},\Phi) \times \mathcal{L}_i^{\forall,X,\ell,A}(\Phi) \times \{i\} \to \mathscr{S}(\{i\},\Phi)$ *is an* acceptable transition rule for a single agent *if for all* $(M,s) \in \mathscr{S}(\{i\},\Phi)$, $\varphi \in \mathcal{L}_i^{\forall,X,\ell,A}(\Phi)$ *if* $\text{PROP}(\varphi) \in \mathcal{L}(s)$, $M = (S,\pi,\mathcal{K}_i,\mathcal{A}_i,\mathscr{PR}_i,\mathcal{L})$, $\tau((M,s),\varphi,i) = (M^*,s^*)$, *and* $M^* = (S^*,\pi^*,\mathcal{K}_1^*,\mathcal{A}_1^*,\mathscr{PR}_1^*,\mathcal{L}^*)$, *then either* $\text{PROP}(\varphi) \subseteq \mathcal{A}_i(s)$ *and* $(M,s) = (M^*,s^*)$, *or* $\text{PROP}(\varphi) \not\subseteq \mathcal{A}_i(s)$ *and there exists a relation* $T \subseteq S \times S^* \times RS(\varphi)$ *such that the following hold:*

T1. *For all* $t^* \in S^*$, *there exists a unique* $t \in S$ *and* f *such that* $(t,t^*,f) \in T$; *moreover,* t^* *is an* f*-replacement of* t.

T2. $(s,s^*,id) \in T$ *and* $\mathcal{A}_i^*(s^*) = \mathcal{A}_i(s) \cup \text{PROP}(\varphi)$.

T3. *If* $\mathscr{PR}_i(s)(\text{CONS}(M,\varphi,i)) \neq 0$, *then for all* $(t,t^*,f) \in T$ *the following conditions hold:*

(a) $\mathcal{K}_i^*(t^*) = \{t^\dagger : (t',t^\dagger,f') \in T$, *for some* $t' \in \mathcal{K}_i(s) \cap \text{CONS}(M,\varphi,i), f' \in RS(\varphi,i,s)\}$.

(b) *If* $t' \in \mathcal{K}_i(s) \cap \text{CONS}(M,\varphi,i)$, *let* $rep(t') = \{t^\dagger \in \mathcal{K}_i^*(t') : \exists f'((t',t^\dagger,f') \in T)\}$. *Then* $\mathscr{PR}_i^*(t^*)(rep(t')) = \mathscr{PR}_i(s)(t')/\mathscr{PR}_i(s)(\text{CONS}(M,\varphi,i))$.

Note that the conditions in the definition apply only if $\text{PROP}(\varphi) \subseteq \mathcal{L}(s)$. State s cannot be the actual world if its language does not include propositions that appear in a formula that is part of the description of the world. We do not care what τ does at situations (M,s) that do not describe the world. For situations that describe the world, nothing changes if i was already aware of all the primitive propositions in φ. If i was not aware of all the primitive propositions in φ, then we use the relation T to talk about corresponding states. T1–T3 describe the key properties of the T relation. T1 says that, in M^*, each state t^* is an f-replacement of some state t in S. T2 ensures that the distinguished state s^* comes from s and, agrees with s as far as the language and truth of primitive propositions goes (since it is an id-replacement of s); moreover, i's awareness changes appropriately (since i becomes aware of φ).

The most interesting requirement is T3, which captures how i's beliefs change in states in $\mathcal{K}_i^*(s^*)$. Notice that, by T2 and the constraints on awareness sets, $\mathcal{A}_i^*(t^*) = \mathcal{A}_i(s) \cup \text{PROP}(\varphi)$ for all states that i considers possible at s^*. Thus, i's awareness must be updated by adding the new propositions in φ, and he knows this has occurred. There are no further constraints if $\mathscr{PR}_i(s)(\text{CONS}(M,\varphi,i)) = 0$; we have nothing to say about how i's beliefs change if i ascribes probability 0 at state s to the possibility of becoming aware of the new primitive propositions in φ.

If i ascribed positive probability to $\text{CONS}(M,\varphi,i)$, then each state $t' \in K_i(s) \cap \text{CONS}(M,\varphi,i)$ corresponds to some states $t^\dagger$ in $\mathcal{K}_i(s^*)$; each such state $t^\dagger$ is a possible replacement of t, for some $f \in RS(\varphi,i,s)$, so f is compatible with i becoming aware of φ at s.

The probability of this set of replacements of t', which we denote $rep(t')$, according to $\mathscr{PR}_i^*(t^*)$ (which is the same as $\mathscr{PR}_i^*(s^*)$) is exactly the probability of t' according to $\mathscr{PR}_i(s)$, conditional on $\text{CONS}(M,i,\varphi)$. Thus, i's beliefs about propositions he becomes aware of depend on his interpretation of the shadow propositions. This updating rule is explored in a probability theoretic setting by (Piermont 2019), where the primitive is a prior and a posterior measure, $\mathscr{PR}_i$ and $\mathscr{PR}_i^*$, and where $\mathscr{PR}_i^*$ can be defined over a possibly richer algebra of events. Piermont provides conditions on this pair of measures that suffice to ensure that they arise via an updating rule as given by T3(c).

The following example explores the mechanics of the single agent transition rule and is presented diagrammatically in Figure 1.

Example 2. *Consider an agent* i *who is uncertain how to interpret a novel proposition. Let* M *be the initial model with a state space* $S = \{s,t\}$, *where* $\mathcal{L}(s) = \{p\}$ *and* $\mathcal{L}(t) = \{q,q'\}$. *Think of* p *as a real proposition and* q *and* q' *as shadow propositions. In* s, p *is true, and in* t, q *is true and* q' *false. The agent is unaware of* p *and considers only* t *possible:* $\mathcal{A}_i(s) = \mathcal{A}_i(t) = \emptyset$ *and* $\mathcal{K}_i(s) = \mathcal{K}_i(t) = \{t\}$. *Obviously, the agent places probability 1 on* t.

For some $\alpha \in [0,1]$, *consider the transition rule* τ^α *such that* $\tau^\alpha((M,s),p,i) = (M^*,s^*)$. M^* *has the state space* $S^* = \{s^*,t^*,t^\dagger\}$. *Let* f *be the* p*-replacement scheme given by* $f : p \mapsto q$ *and let* f' *be the* p*-replacement scheme given by* $f' : p \mapsto q'$. *Then* $T = \{(s,s^*,id),(t,t^*,f),(t,t^\dagger,f')\}$. *In all states,* i*'s awareness is exactly* p, *and in all states, she considers both* t^* $t^\dagger$ *possible, that is,* $\mathcal{K}_i(s^*) = \mathcal{K}_i(t^*) = \mathcal{K}_i(t^\dagger) = \{t^*,t^\dagger\}$. *Finally,* i *puts probability* α *on state* t^* *and* $(1-\alpha)$ *on state* $t^\dagger$. *When* $\alpha = 1$, i *is sure that the novel proposition* p *is what she was representing by the shadow proposition* q, *and when* $\alpha = 0$ *she is sure it was represented by* q'. *In between, the agent is uncertain about the interpretation of the novel proposition. Notice this uncertainty is not captured in* M, *as the agent is unaware of* p *in* M, *hence can not directly reason about its likelihood of being represented by* q *or* q'. ∎

3.2 Dynamics: The Multiagent Case

In this section, we present a transition rule for multiagent models. We assume that other agents do not realize that i has become aware, so their beliefs remain invariant under the model transition. Although this is conceptually simple, it complicates the definition of an acceptable rule because the updated model must contain additional states to handle other agents' beliefs. For example, if there are two agents, and j initially (correctly) believes i is unaware of p, then after i becomes aware of p, j must consider a state where i is unaware of p, even though he no longer is. Furthermore, since j still believes i is unaware of p, the worlds that i considers from the worlds that j considers (i.e., $\mathcal{K}_i(\mathcal{K}_j(s))$) cannot contain p in language, requiring yet more states to be added.

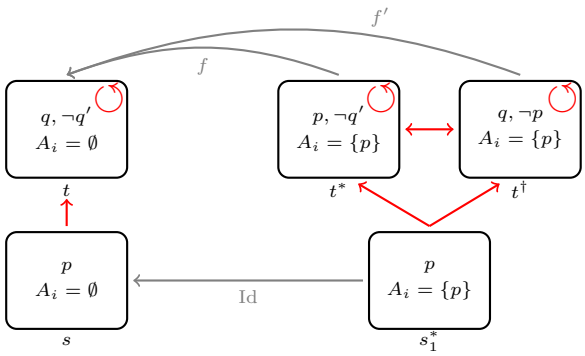

Figure 1: A visual representation of Example 2. The states on the left are the initial state space, M, and those on the right are M^*. The red, unlabeled arrows indicate i's accessibility relation. The gray arrows indicate the relation T, labeled according to the replacement scheme.

The following definition generalizes Definition 2 to models with multiple agents.

Definition 3. *$\tau : \mathscr{S}(\mathcal{I}, \Phi) \times \mathcal{L}_n^{\forall, X, \ell, A}(\Phi) \times \mathcal{I} \to \mathscr{S}(\mathcal{I}, \Phi)$ is an acceptable transition rule if for all $(M, s) \in \mathscr{S}(\mathcal{I}, \Phi)$, $\varphi \in \mathcal{L}_n^{\forall, X, \ell, A}(\Phi)$, and $i \in \mathcal{I}$, if $\text{PROP}(\varphi) \in \mathcal{L}(s)$, $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \mathscr{PR}_1, \ldots, \mathscr{PR}_n, \mathcal{L})$, $\tau((M, s), \varphi, i) = (M^*, s^*)$, and $M^* = (S^*, \pi^*, \mathcal{K}_1^*, \ldots, \mathcal{K}_n^*, \mathcal{A}_1^*, \ldots, \mathcal{A}_n^*, \mathscr{PR}_1^*, \ldots, \mathscr{PR}_n^*, \mathcal{L}^*)$, then either $\text{PROP}(\varphi) \subseteq \mathcal{A}_i(s)$ and $(M, s) = (M^*, s^*)$, or $\text{PROP}(\varphi) \not\subseteq \mathcal{A}_i(s)$ and there exists a relation $T \subseteq S \times S^* \times RS(\varphi)$ such that the following hold:*

T1. *For all $t^* \in S^*$, there exists a unique $t \in S$ and f such that $(t, t^*, f) \in T$; moreover, t^* is an f-replacement of t.*

T2. *$(s, s^*, id) \in T$, $\mathcal{A}_i^*(s^*) = \mathcal{A}_i(s) \cup \text{PROP}(\varphi)$, and $\mathcal{A}_j^*(s^*) = \mathcal{A}_j(s)$ for $j \neq i$.*

T3. *If $(t, t^*, f) \in T$, $t \in \mathcal{K}_i(s) \cup \{s\}$, either $f \neq id$ or $(t^*, f) = (s^*, id)$, and $\mathscr{PR}_i(s)(\text{CONS}(M, \varphi, i)) \neq 0$, then the following conditions hold:*

(a) *$\mathcal{K}_i^*(t^*) = \{t^\dagger : (t', t^\dagger, f') \in T$, for some $t' \in \mathcal{K}_i(s) \cap \text{CONS}(M, \varphi, i), f' \in RS(\varphi, i, s)\}$.*

(b) *If $t' \in \mathcal{K}_i(s) \cap \text{CONS}(M, \varphi, i)$, let $rep(t') = \{t^\dagger \in \mathcal{K}_i^*(t') : \exists f'((t', t^\dagger, f') \in T)\}$. Then $\mathscr{PR}_i^*(t^*)(rep(t')) = \mathscr{PR}_i(s)(t') / \mathscr{PR}_i(s)(\text{CONS}(M, \varphi, i))$.*

(c) *$f(\mathcal{A}_j(t^*)) = \mathcal{A}_j(t)$ for all $j \neq i$.*

T4. *If $(t, t^*, f) \in T$, then the following hold for all agents $j \neq i$, and for $j = i$ if $t \notin \mathcal{K}_i(s) \cup \{s\}$ or if $f = id$ and $t^* \neq s^*$:*

(a) *$f(\mathcal{A}_j(t^*)) = \mathcal{A}_j(t)$.*

(b) *For all $t' \in \mathcal{K}_j(t)$, there exists $t^\dagger \neq s^*$ such that $(t', t^\dagger, f) \in T$, and $\mathcal{K}_j^*(t^*) = \{t^\dagger : (t', t^\dagger, f) \in T$ for some $t' \in \mathcal{K}_j(t)$, $t^\dagger \neq s^*\}$.*

(c) *$\mathscr{PR}_j^*(t^*)(rep(t') \cap \mathcal{K}_i(t^*)) = \mathscr{PR}_j(s)(t')$.*

T1 and T2 are the same as in Definition 2 with the caveat that the awareness of agents other than i does not change. T3 is largely the same as before, except that it adds a requirement handling the awareness of other agents and it does not apply to all states. For a state $t^\dagger$ that corresponds to t' via f, the awareness of each agent $j \neq i$ changes according to f; if i becomes aware of p, then i replaces $f(p)$ with p, so if j was aware of $f(p)$ in t', then j must be aware of p in $t^\dagger$—p replaces $f(p)$ in $\mathcal{A}_j(t^*)$.

The reason T3 no longer applies to all states is we need to allow other agents to maintain their beliefs. Note that states $t^\dagger$ for which $(t', t^\dagger, id) \in T$ are *not* in $\mathcal{K}_i(s^*)$, since $id \notin RS(\varphi, i, s)$ if there propositions that i is not aware of in φ and thus are not under the scope of T3. This is because $\text{PROP}(\varphi) \subseteq \Phi$, which is disjoint from Φ', so we cannot have $id(\text{PROP}(\varphi) - \mathcal{A}_i(s)) \subseteq \Phi'$. We use states $t^\dagger$ where $(t', t^\dagger, id) \in T$ and $t' \in \mathcal{K}_i(s)$ to capture other agents' beliefs about i.

Finally, T4 describes agents' beliefs and awareness except for i's belief and awareness in states in $\mathcal{K}_i(s^*)$. Because other agents do not know about i's increased awareness, T4 also handles the construction of i's beliefs and awareness in states that are considered possible by other agents. Roughly speaking, T4 says that if $(t, t^*, f) \in T$, then each agent's awareness changes according to f. The states that j considers possible at a state that is an f-replacement of t are exactly the f-replacements of $\mathcal{K}_j(t)$. So knowledge is unperturbed except as required by replacing shadow propositions where needed. The final condition of T4 mirrors this but for probabilistic assessments.

Example 3. *Consider what happens when agent i in the model M from Example 1 becomes aware of p' in state s_1. Let τ be a transition rule that maps (M, s_1) to (M^*, s_1^*) with states $S^* = \{s_1^*, s_{2i}^*, s_{1j}^*, s_{2j}^*, s_{3j}^*, s_{2ij}^*, s_{3ij}^*, s_{1ij}^*\}$. Letting f denote the p'-replacement that maps p' to q, T consists of (s_1, s_1^*, id), (s_2, s_{2i}^*, f), and (s_n, s_{nj}^*, id) and $(s_n, s_{nij}^*, f)\}$ for $n \in \{1, 2, 3\}$.*

The awareness functions in M^* are defined as follows: $\mathcal{A}_i^*(s_1^*) = \mathcal{A}_j^*(s_1^*) = \mathcal{A}_i^*(s_{2i}^*) = \mathcal{A}_j^*(s_{2i}^*) =$

$\mathcal{A}^*_j(s^*_{2j}) = \{p, p'\}$ *and* $\mathcal{A}^*_i(s^*_{2j}) = \mathcal{A}^*_i(s^*_3) = \mathcal{A}^*_j(s^*_3) = \{p\}$. *The reachability functions in* M^* *are defined as follows:* $\mathcal{K}^*_i(s^*_1) = \mathcal{K}^*_i(s^*_{2i}) = \{s^*_{2i}\}$; $\mathcal{K}^*_j(s^*_1) = \mathcal{K}^*_j(s^*_{2i}) = \mathcal{K}^*_j(s^*_{2j}) = \{s^*_{2j}\}$; $\mathcal{K}^*_i(s^*_{2j}) = \mathcal{K}^*_i(s^*_3) = \{s^*_{2j}, s^*_3\}$; *and* $\mathcal{K}^*_j(s^*_3) = \{s^*_3\}$.

At s^*_1, *i believes that he is fully aware and also knows that p is true. This is because he has disregarded the possibility of* s_3, *since* $s_3 \notin \text{CONS}(p', i)$. *Thus, despite fact that i only became aware of* p', *and did not directly learn the truth of any proposition, he has implicitly learned that p holds. Moreover, notice that j believes i is unaware of* p'. ∎

The transition rule in Example 3, as opposed to that in Example 2, adds many more states. This is because when agent i becomes aware of p', agent j does not know this happened. Thus, we need a set of states that represent j's (now incorrect) beliefs that i is unaware of p': these are the states s^*_{1j}, s^*_{2j}, and s^*_{3j}. In addition, i knows that j does not know she became aware of p'. So we also have to add more states to capture i's correct beliefs about j's incorrect beliefs about i's awareness; these are the states s^*_{1ij}, s^*_{2ij}, and s^*_{3ij}.

3.3 Dynamics: Syntax and Axioms

In this section we briefly explore an extension of the language presented in Section 2 that allows us to capture the dynamic aspects of our transition rule. We then present some axioms which we hope will help convey the requirements of a valid transition rule more directly. While these axioms are all sound, getting a complete axiomatization seems to require more effort. We are currently exploring this.

Given that we are interested in what happens after an agent becomes of a formula, it seems reasonable to consider extending the language by adding formulas of the form $[\varphi, i]\psi$ to the underlying language $\mathcal{L}^{\forall,X,\ell,A}_n$ considered in this paper, where $[\varphi, i]\psi$ is interpreted as "after i becomes aware of φ, ψ is true." Such a formula is interpreted relative to a probabiistic awarness structure M and a transition rule τ in the obvious way:

$$(M, s, \tau) \models [\varphi, i]\psi \text{ if } ((\tau((M, s), \varphi, i), \tau) \models \psi.$$

Note that this definition allows for occurrences of formulas of the form $[\varphi', j]\psi'$ in ψ (although not in φ).

In this language, we can capture many of the properties of the transition function τ. For example, the following two axioms are sound:

A*. $A_i\psi \Leftrightarrow [\varphi, i]A_i(\varphi \wedge \psi)$ if $\text{PROP}(\psi) \cap \text{PROP}(\varphi) = \emptyset$.

AKo. $K_j\psi \wedge A_j\psi \Leftrightarrow [\varphi, i](K_j\psi \wedge [\varphi, i]A_j\psi)$.

A* states that after becoming aware of φ agent i is aware of φ and everything that he was initially aware of. AKo states that the knowledge and awareness of other agents remains invariant.

To capture how an agent's beliefs changes as a result of an update, we need to be able to talk about consistency. Let $Consis(\varphi, i)$ be a new formula that is true at a state t if t is consistent with i becoming aware of φ. If $\text{PROP}(\varphi)$ has no primitive propositions, then $Consis(\varphi, i)$ is equivalent to *true*, and if $|\text{PROP}(\varphi)| = 1$, then $Consis(\varphi, i)$ is equivalent to $\exists x \neg A_i(x)$. But if $|\text{PROP}(\varphi)| > 1$, then it can be shown that $Consis(\varphi, i)$ is not equivalent to a formula in $\mathcal{L}^{\forall,X,\ell,A}_n$. Using $Consis(\varphi, i)$, we can provide some sound axioms that capture how an agent's beliefs change:

Ka. $(\ell_i(Consis(i, \varphi) > 0) \wedge K_i(Consis(i, \varphi) \Rightarrow \psi)) \Rightarrow [\varphi, i]K_i\psi$.

Kb. $(\ell_i(Consis(i, \varphi) > 0 \wedge [\varphi, i]K_i\psi[x/\psi']) \Rightarrow K_i(Consis(\varphi, i) \Rightarrow \exists x\psi)$.

Pra. $\ell_i(\psi \wedge Consis(\varphi, i)) > \alpha\ell_i(Consis(\varphi, i)) \Rightarrow [\varphi, i](\ell_i(\psi) > \alpha)$.

Prb. $(\ell_i(Consis(i, \varphi) > 0 \wedge [\varphi, i](\ell_i\psi[x/\varphi]) > \alpha) \Rightarrow \ell_i(\exists x\psi \wedge Consis(\varphi, i)) > \alpha\ell_i(Consis(\varphi, i))$.

Ka states that after becoming aware of φ, i knows everything he knew before he was aware of φ that was consistent with becoming aware of φ, provided that becoming aware of φ has positive probability. Kb is almost a converse. It says that if after becoming aware of φ i knows ψ, then before becoming aware, i knew that ψ (with occurrences of φ replaced by an existential) held in all states where it was consistent for i to become aware of φ. Pra and Prb are analogues of Ka and Kb for probability formulas. Note that $\ell_i(\psi \wedge Consis(\varphi, i)) > \alpha\ell_i(Consis(\varphi, i))$ essentially says that the probaiblity of ψ conditional on $Consis(\varphi, i)$ is greater than α. Since we do not have conditional probability formulas in the language, we cannot say this directly.

4 Information Disclosure

The following section is a simple application of our model to the problem of information disclosure. In the interest of space and simplicity, we consider a highly stylized setting: There are two agents, a buyer and a seller (we can think of the seller as a restaurateur and the buyer as a potential patron). The agents' belief and awareness will be modeled via a probabilistic awareness structure for 2 agents along with a transition rule. We can think of the agents as playing a game (i.e., they have actions and payoffs associated with these actions), but the game is not described explicitly in the epistemic model.

The initial state space (i.e., before the buyer become aware of a formula of which he was previously unaware) is $\{s^m_n, t^m_n\}_{(n,m)\in\{1,\ldots,N\}\times\{1,\ldots,M\}}$. The lower index (n) represents the *type* or quality of the seller. So in state s^m_n or t^m_n, the seller has quality n; we assume that the buyer's value for a purchase is increasing in n (i.e., the buyer gets better-quality meals, which he values more highly, at a better-quality restaurant). The upper index (m) represents the buyer's type: how the buyer reacts to becoming aware of new features (explained in detail shortly).

The s states (i.e., states of the form s^m_n for some n and m) are "real" in the sense that they contain no

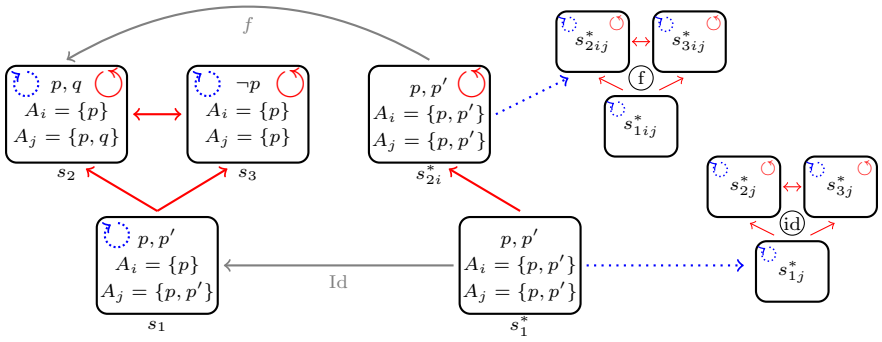

Figure 2: A visual representation of Example 3. The red arrows indicate i's accessibility relation; the blue (dotted) arrows indicate j's accessibility relation. The gray arrows indicate the relation T, labeled according to the replacement scheme. The two sets of three states on the right side of the figure are both copies of the original three states, under different replacement schemes; the upper three states are f-replacements and the lower three are id-replacements.

shadow propositions. The seller, who we assume is fully aware, considers only s states possible. The t states represent the buyer's understanding of the world, containing shadow propositions in place of the real propositions that he is unaware of but exist in the s states. We assume that (it is commonly known that) the seller knows his type n but not how the buyer will react to information, so $\mathcal{K}_s(s^m_n) \subseteq \{s^{m'}_n : m' \in \{1, \ldots, M\}\}$. On the other hand, (it is commonly known that) the buyer knows how he will react to information, but does not know the seller's type, so $\mathcal{K}_b(s^m_n) = \mathcal{K}_b(t^m_n) = \{t^m_{n'} : n' \in \{1, \ldots, N\}\}$.

A list of propositions $\{p_1, \ldots, p_K\}$ is a *rating structure*, and the p_k's are called *ratings*, if (1) exactly one proposition p_k is true at each state s^m_n and (2) if p_k is true at s^m_n, $p_{k'}$ is true at $s^{m'}_{n'}$, and $k > k'$, then $n > n'$. Thus, higher ratings indicate higher-quality sellers. Note that this condition implies that the same rating must hold at s^m_n and $s^{m'}_n$: the seller knows the rating. In general, a rating may be true at more than one state; a rating p is *perfectly informative* if there is exactly one n such that p is true of only states of the form s^m_n. We assume that the seller is aware of the rating, but the buyer is not. Formally, we assume a particular rating structure, given by $p_1, \ldots, p_K$, such that the seller is aware of $p_1, \ldots, p_K$ at all states of the form s^m_n. The buyer is not aware of these formulas at any state. In some states of the form t^n_m there is no rating structure in the language (these are the states where it does not occur to the buyer that there is a rating structure). In other states there of the form t^m_n, there are one or more rating structures, but these are represented by shadow propositions $x_1, \ldots, x_{K'}$; the buyer is not aware of these shadow propositions, but understands that the seller knows the actual rating structure (so, in these states, formulas of the form $\neg A_b(x_j) \wedge A_s(x_j) \wedge (K_s(x_j) \vee K_s(\neg x_j)$ hold). We assume that the superscript m in t^m_n indicates which rating structures are in the language at t^m_n, so if $\{x_1, \ldots, x_K\}$ is a rating structure and $\{x_1, \ldots, x_K\} \subseteq \mathcal{L}(t^m_n)$, then $\{x_1 \ldots x_K\} \subseteq \mathcal{L}(t^m_{n'})$ for all n'.

The seller sets a price and possibly discloses his rating; the buyer then decides whether to buy (i.e., whether to eat at the restaurant). Before deciding, the buyer updates his beliefs based on whatever information the seller discloses and his beliefs about the seller's. If the buyer was unaware of p_k and p_k is disclosed, then the buyer first updates his beliefs according to the transition rule to become aware of p_k, then he conditions on the fact that p_k is true. Among other things, the index m captures how the buyer will update when she becomes aware of a rating (i.e., it encodes a replacement function). An equilibrium of this game consists of (1) an action for each type of seller, and (2) beliefs and an action for the buyer, for each action of the seller, where (1) and (2) are such that both parties are maximizing (the expected revenue and the expected value of the purchase, respectively).

As we observed in the introduction, if the buyer is aware of the rating structure, then all sellers (except possibly those with the lowest rating) will reveal their rating. On the other hand, if the buyer is unaware of the rating structure $\{p_1, \ldots, p_K\}$, we must explicitly indicate how he would treat the lack of revelation. This is the role of the buyer's type. We assume here that when no rating is revealed, the buyer's beliefs do not change at all: the buyer continues to be unaware of the rating. But note that this is fundamentally different from the buyer knowing that the rating could have been disclosed but was not. Under this assumption, the unraveling argument given in the introduction fails: any type with a rating imparting lower-than-average quality would prefer the buyer to retain his original belief; not disclosing does not have any effect on these beliefs, since the buyer's unawareness prevents him from reasoning about the actions of other types.

Given the immediate failure of the unraveling argument, the more interesting aspect of this model is the dependency of the seller's beliefs on the *transition rule*,

as encoded by the m index. The key point is that s_n^m and $s_n^{m'}$ are distinguished by the buyer's reaction to becoming aware of a new rating. We provide an example showing how the seller's beliefs about the transition rule determine whether the seller decides to disclose his rating.

Suppose that $N = 3$, so that there are three qualities of seller (restaurant) and $M = 2$. Suppose that the buyer's utility for a purchase if the restaurant has quality n is n, and that at s_n^m, the buyer initially believes that t_1^m, t_2^m, and t_3^m have probability $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{4}$, respectively. This indicates that in the absence of any signal, when the buyer does not update his beliefs at all, his expected utility is $1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} = \frac{7}{4}$.

Suppose that the rating p_n is true at state s_n^m, for $n = 1, 2, 3$, so p_n is perfectly informative; $\mathcal{L}(s_n^m) = \{p_1, p_2, p_3\}$. We take $\mathcal{L}(t_m^n) = \{x_1, x_2, x_3, y_1\}$, where these are all shadow ratings. Think of x_n as the shadow signal corresponding to p_n, so that x_n is true only at the states t_n^m for $n = 1, 2, 3$. The shadow rating y_1 is a coarsening of x_1 and x_2; it is true at both t_1^m and t_2^m.

The transition rule is such that, if the buyer becomes aware of p_n, the situation (M, s_n^m) is mapped to the situation $(M_n^*, s_n^{m^*})$ with relation T_n containing $(t_{n'}^1, t_{n'}^{1*}, f_n^1)$ and $(t_{n'}^2, t_{n'}^{2*}, f_n^2)$ for $n' \in \{1, 2, 3\}$, where the replacements are given by $f_n^1 : p_n \mapsto x_n$ for $n \in \{1, 2, 3\}$, and $f_1^2 : p_1 \mapsto y_1$, $f_2^2 : p_2 \mapsto y_1$, and $f_3^2 : p_3 \mapsto x_3$. That is, in states of the form t_n^1, the buyer interprets p_k as arising from the perfectly informative shadow rating structure $\{x_1, x_2, x_3\}$, and in states t_n^2, he interprets it as arising from the coarser shadow rating structure $\{y_1, x_3\}$, conflating p_1 and p_2.

As always, the highest type, a seller in state s_3^m, will disclose regardless of his beliefs about the transition rule. Conversely, a seller in state s_1^m never has a reason to disclose. However, a seller in state s_2^m will disclose only when he is sufficiently confident the seller will correctly interpret p_2 as being perfectly informative. In state s_2^1, disclosing imparts beliefs (in the updated model) that places probability 1 on state t_2^1, and a willingness to pay of $2 > \frac{7}{4}$. Thus, if he knew that he was in state s_2^1, he would certainly want to disclose p_2. If, however, he knew he was in state s_2^2, disclosing imparts a beliefs (in the updated model) that places probability $\frac{2}{3}$ on state t_1^1 and $\frac{1}{3}$ on state t_2^1, and a willingness to pay of $\frac{4}{3} < \frac{7}{4}$.

5 Discussion

This paper develops a modal logic that allows agents to reason both about probabilities and about their own and others' awareness. We introduce the semantic notion of a model transition, which captures how a model changes when an agent becomes aware of a new formula. In contrast to prior work on the subject, our transition rule allows the relative likelihood of two formulas that an agent was aware of to change arbitrarily after she becomes more aware. This is because the agent can, to some extent, reason about her own unawareness. If, for example, after becoming aware of φ she believes that having been unaware of φ was correlated with the truth of some other formula ψ, then her probabilistic assessment of ψ can change.

We then apply this model to a simple economic environment of information disclosure, where one agent can strategically decide to make another agent aware of a new formula. We show that the agent's beliefs about how others will react to novel information (i.e., her beliefs about the model transition rule) determine her decision to disclose information and expand the awareness of other agents.

In future work, we hope to explore further applications of dynamic awareness. We believe that thinking about how awareness changes will prove critical in understanding economic puzzles and the general problem of updating in the presence of unawareness. We would also hope to provide a sound and complete axiomatic characterization of our transition rule, in the spirit of the discussion in Section 3.3.

Acknowledgments: Halpern was supported in part by NSF grants IIS-178108 and IIS-1703846, a grant from the Open Philanthropy Foundation, ARO grant W911NF-17-1-0592, and MURI grant W911NF-19-1-0217.

References

O. Board and K.-S. Chung. Object-based unawareness: theory and applications. Working paper 378, University of Pittsburgh, 2009.

J. C. Cox, M. Servátka, and R. Vadovič. Status quo effects in fairness games: reciprocal responses to acts of commission versus acts of omission. *Experimental Economics*, 20(1):1–18, 2017.

I. Esponda and E. Vespa. Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202, 2014.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41(2):340–367, 1994.

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, MA, 1995. A slightly revised paperback version was published in 2003.

J. Y. Halpern and L. C. Rêgo. Reasoning about knowledge of unawareness revisited. *Mathematical Social Sciences*, 66(2):73–84, 2013.

A. Heifetz, M. Meier, and B. Schipper. Interactive unawareness. *Journal of Economic Theory*, 130:78–94, 2006.

G. Z. Jin, M. Luca, and D. Martin. Is no news (perceived as) bad news? An experimental investigation of information disclosure. Technical report, National Bureau of Economic Research, 2018.

L. K. John, K. Barasz, and M. I. Norton. Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences*, 113(4):954–959, 2016.

E. Karni and M.-L. Vierø. "Reverse Bayesianism": A choice-based theory of growing awareness. *American Economic Review*, 103(7):2790–2810, 2013.

S. Modica and A. Rustichini. Awareness and partitional information structures. *Theory and Decision*, 37:107–124, 1994.

S. Modica and A. Rustichini. Unawareness and partitional information structures. *Games and Economic Behavior*, 27(2):265–298, 1999.

E. Piermont. Unforeseen evidence. *arXiv preprint arXiv:1907.07019*, 2019.

G. Sillari. Quantified logic of awareness and impossible possible worlds. *Review of Symbolic Logic*, 1(4):514–529, 2008.