

Belardinelli, Gaia; Schipper, Burkhard C.

**Working Paper**

## Implicit Knowledge in Unawareness Structures

Working Paper, No. 360

**Provided in Cooperation with:**

University of California Davis, Department of Economics

*Suggested Citation:* Belardinelli, Gaia; Schipper, Burkhard C. (2023) : Implicit Knowledge in Unawareness Structures, Working Paper, No. 360, University of California, Department of Economics, Davis, CA

This Version is available at:

<https://hdl.handle.net/10419/283202>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# IMPLICIT KNOWLEDGE IN UNAWARENESS STRUCTURES\*

Gaia Belardinelli<sup>†</sup>      Burkhard C. Schipper<sup>‡</sup>

October 27, 2023

## Abstract

Awareness structures by Fagin and Halpern (1988) (FH) feature a syntactic awareness correspondence and accessibility relations modeling implicit knowledge. They are a flexible model of unawareness, and best interpreted from an outside modeler’s perspective. Unawareness structures by Heifetz, Meier, and Schipper (2006, 2008) (HMS) model awareness by a lattice of state spaces and explicit knowledge via possibility correspondences. Sublattices thereof can be interpreted as subjective views of agents. Open questions include (1) how implicit knowledge can be defined in HMS structures, and (2) in which way FH structures can be extended to model the agents’ subjective views. In this paper, we address (1) by defining implicit knowledge such that it is consistent with explicit knowledge in HMS models. We also introduce a variant of HMS models that instead of explicit knowledge, takes implicit knowledge and awareness as primitives. Further, we address (2) by introducing a category of FH models that are modally equivalent relative to sublanguages and can be interpreted as agents’ subjective views depending on their awareness. These constructions allow us to show an equivalence between HMS and FH models. As a corollary, we obtain soundness and completeness of HMS models with respect to the Logic of Propositional Awareness, based on a language featuring *both* implicit and explicit knowledge.

**Keywords:** Unawareness, awareness, implicit knowledge, explicit knowledge.

**JEL-Classifications:** D83, C70.

---

\*We thank three anonymous reviewers for TARK 2023, Hans van Ditmarsch, participants in TARK 2023 at Oxford University, CSLI Workshop 2022 at Stanford University, and the 2022 Zoom Mini-Workshop on Epistemic Logic with Unawareness for helpful comments. An extended abstract of the paper has been published in the proceedings of TARK 2023. Gaia gratefully acknowledges funding from the Carlsberg Foundation. Burkhard gratefully acknowledges financial support via ARO Contract W911NF2210282.

<sup>†</sup>Center for Information and Bubble Studies, University of Copenhagen. Email: belardinelli@hum.ku.dk

<sup>‡</sup>Department of Economics, University of California, Davis. Email: bcschipper@ucdavis.edu

# 1 Introduction

Models of unawareness are of interest in various disciplines, most notably in computer science, economics, game theory, decision theory, and philosophy. The seminal contribution in computer science and philosophy are awareness structures by Fagin and Halpern (1988) (henceforth, *FH models*) who extended Kripke structures with a syntactic awareness correspondence in order to feature notions of implicit knowledge, explicit knowledge, and awareness. In economics, Heifetz, Meier, and Schipper (2006, 2008) introduced unawareness structures (henceforth, *HMS models*) that consist of a lattice of state spaces featuring a notion of explicit knowledge and awareness. Like Kripke structures, HMS models can be constructed canonically and three different sound and complete axiomatizations have been presented (Halpern and Rêgo, 2008, Heifetz, Meier, and Schipper, 2008).<sup>1</sup> There have already been a number of applications to game theory, decision theory, mechanism design and contracting, financial markets, electoral campaigning, conflict resolution, social network formation, business strategy and entrepreneurship etc.<sup>2</sup>

The different modeling approaches may be seen as reflecting the different foci of the fields. HMS models in economics are very much motivated by game theory and its applications. The main underlying idea is that *explicit* rational reasoning of players is what drives their behavior. Hence, the model features only explicit knowledge (without the detour via implicit knowledge) and awareness, and it can be interpreted as encompassing the subjective views of players.<sup>3</sup> Moreover, the syntax-free frame lends itself seamlessly to the existing body of work in decision theory and game theory.<sup>4</sup> The focus on behavioral implications also explains why the model is built on strong properties of knowledge such as (positive) introspection and factivity: this allows for the identification of the behavioral implications of unawareness per se without confounding it with mistakes in information processing. Coming from a different angle, FH models were motivated more generally by the study of the logical non-omniscience problem in computer science and philosophy (see e.g., Hintikka, 1975, Levesque, 1984, Lakemeyer, 1986, Stalnaker, 1991). They represent awareness via syntactic awareness correspondences, which for each agent

---

<sup>1</sup>For other approaches and an overview, see Schipper (2015).

<sup>2</sup>*Game theory*: See for instance, Rêgo and Halpern (2012), Heifetz, Meier, and Schipper (2013b, 2021), Grant and Quiggin (2013), Halpern and Rêgo (2014), Meier and Schipper (2014a, 2023), Guarino (2020), Feinberg (2021), Schipper (2021), and Perea (2022). *Decision theory*: See for instance, Karni and Vierø (2013), Schipper (2013, 2014), and Dominiak and Tserenjigmid (2018). *Contract theory and mechanism design*: See for instance, von Thadden and Zhao (2012), Filiz-Ozbay (2012), Auster (2013), Chung and Fortnow (2016), Auster and Pavoni (2021a), Francetich and Schipper (2021), and Pram and Schipper (2023). *Financial markets*: See Heifetz, Meier, and Schipper (2013a), Meier and Schipper (2014b), Galanis (2018), Auster and Pavoni (2021b), Schipper and Zhou (2021). *Electoral campaigning*: Schipper and Woo (2019). *Conflict resolution*: Rêgo and Vieira (2020). *Social network formation*: Schipper (2016). *Business strategy*: Bryan, Ryall, and Schipper (2021).

<sup>3</sup>These subjective views of agents are further elucidated by Grant et al. (2015).

<sup>4</sup>For example, Schipper (2014), Heifetz, Meier, and Schipper (2013a), Meier and Schipper (2014a), Rêgo and Vieira (2020), Schipper and Zhou (2021), Pram and Schipper (2023).

assigns a set of formulas to each state. This approach to awareness modeling offers a great deal of flexibility, because the set of formulas an agent is aware of may be arbitrary, thereby allowing potentially for the representation of different notions of awareness.<sup>5</sup> However, because their semantics is not syntax-free, their applications to decision or game theory require more effort. This is because in decision theory and game theory and applications thereof, the primitives are typically not described syntactically. Moreover, FH models are best interpreted as a tool used by an outside modeler for two reasons: First, the primitive notion of knowledge is *implicit* knowledge while explicit knowledge is derived from implicit knowledge and awareness. Implicit knowledge is not necessarily something that the agent herself can consciously reason about. Second, we cannot think of FH models as models that the agents themselves use for analyzing their epistemic universe unless they are aware of everything. Rather, FH models capture the perspective of a systems designer of a multi-agent distributed system. This becomes relevant in interactive settings when we are interested in the players' interactive perceptions of the epistemic universe. Despite the differences in modeling approaches, Halpern and Rêgo (2008) and Belardinelli and Rendsvig (2022) formalize in which ways HMS models are equivalent to FH models in terms of explicit knowledge and awareness. However, as the discussion above makes clear, there remain open questions: First, can implicit knowledge be captured also in HMS models and how would this notion of implicit knowledge be related to implicit knowledge in FH models? Second, can we extend FH models so as to interpret them from the agents' subjective point of views? These questions will be answered in this paper.

The prior discussion begs a more fundamental question: Why would it even be meaningful to provide answers to the two above questions given the differences in the modeling philosophies between computer science and economics? From a philosophical standpoint, it is especially interesting to note that there is a notion of knowledge, namely implicit knowledge, that seems to be eschewed in one field but accepted essentially without much discussion in another. Since the study of awareness is by and large interdisciplinary, we think it is paramount to have a multidisciplinary dialogue exploring the differences between fields addressing the topic and perhaps even resolving these differences. This motivated us to present, as a prelude to our formal study, a literal but fictive dialogue between various disciplines involved in the study of awareness. Envision an economist (E), computer scientist (C), psychologist (P), neuroscientist (N), and sociologist (S) are at dinner together in a restaurant. The economist and computer scientist start a conversation, while others initially chat on a different topic.

- E     I don't understand implicit knowledge in the context of awareness...
- C     What's the matter? Implicit knowledge begets explicit knowledge when raised into awareness.
- E     Yes, I know that you define explicit knowledge as implicit knowledge and awareness, but what is implicit knowledge?

---

<sup>5</sup>See Fagin and Halpern (1988, pp. 54-55) and Fagin et al. (1995, Chapter 9.5) for discussions.

- C** Well, there is nothing mysterious about it. In economics, you are used to model knowledge with partitions in Aumann structures or, as we call it in computer science and logic, indistinguishability relations in Kripke frames. In FH models, implicit knowledge is what is modeled by partitions or more generally accessibility relations.
- E** Implicit knowledge in FH models looks formally similar to knowledge in Kripke or Aumann structures. However, I think the connotation is quite different. In Aumann or Kripke structures, partitions model explicit knowledge. It is knowledge that the agent herself can consciously reason about rather than just being ascribed to her by the modeler.
- C** That's because in Aumann and Kripke structures there is no unawareness.
- E** Precisely. Partitions model explicit knowledge by construction.
- C** FH models offer more "structure" than Kripke models, which allows us to distinguish between implicit and explicit knowledge. Explicit knowledge is derived from implicit knowledge via awareness. Implicit knowledge is the more primitive concept and thus naturally modeled by accessibility relations.
- E** Why is implicit knowledge an appropriate primitive of a theory of knowledge under unawareness? Explicit knowledge is what a decision maker herself can consciously reason about, what affects her decision-making, and thus becomes testable in behavioral data. Why would we want a theory that is based on the untestable construct of implicit knowledge? There is no need to take it as the primitive. As HMS models demonstrate, we can easily take explicit knowledge as the primitive. We can have a testable theory of unawareness.
- C** Hold on, hold on. You seem to be thinking about awareness and knowledge only in relation to *human* decision making, but there are other important contexts as well. For instance, we are very interested in the knowledge of distributed systems.<sup>6</sup> The system designer may have a pretty good idea of the knowledge implicit in the system and just needs a model to represent it.
- E** Your focus on knowledge systems is interesting. It explains the outside modeler's perspective implicit in FH models.<sup>7</sup> You logicians are quick to endow agents with a formal language but when it comes to models, you keep them stingily for yourself. It is as if all agents that you model live in the pre-Kripke age. That is, you rarely allow your agents to reason within the model itself; models are used by modelers only! Take FH models, could agents use them to analyze their own situation? No, unless everybody is aware of everything.

---

<sup>6</sup>Fagin, Halpern, and Vardi (1986) and Halpern (1987).

<sup>7</sup>This point is made eloquently by Fagin, Halpern, and Vardi (1986): "The notion of knowledge is *external*. A process cannot answer questions based on its knowledge with respect to this notion of knowledge. Rather, this is a notion meant to be used by the system designer reasoning about the system. ... (I)t does seem to capture the type of intuitive reasoning that goes on by system designers."

- C** Why would agents need models? They reason in a language. Models are for the systems designer to analyze the system.
- E** Well, in game theory we are interested in how the subjective interpretations of a situation drives the agents' interactive behavior. Standard game theoretic models may be thought of as being shared among players and the modeler. Any player could be a modeler who models various uncertainties and all players' possible state of minds.<sup>8</sup> In games with unawareness using HMS models, we allow each player to be a modeler using her sublattice of spaces of the lattice.<sup>9</sup> Typically, we only introduce a formal language to investigate the logical foundations of game theoretic models.
- C** Game theoretic modeling is done mainly by analysts. I do not see ordinary people writing game theoretic models when involved in interactions. The lack of subjective interpretation of FH models seems to be just a minor aesthetic issue to me that is of little practical consequence. I also believe we could easily define subjective versions of FH models. Any subjective version of a FH model should be bisimilar to the original one, modulo awareness. Like HMS models, the set of bisimilar FH models would constitute a lattice by set inclusion of atomic formulas.
- E** I would like to see how exactly you construct your subjective versions of FH models. However, the fact that FH models do not have a syntax-free semantics would make them impractical for modeling game theoretic applications.
- C** This seems to be a problem of economists. Working with syntax is very natural in computer science. Again, the lack of syntax-freeness of the semantics of FH models is a minor aesthetic issue that is well overcompensated by the flexibility they provide in modeling various notions of unawareness. And the multiple state spaces of HMS models should make them also more challenging to apply than let's say Kripke frames or Aumann models, shouldn't they?
- E** Fair enough. Let's come back to my question: What really is implicit knowledge in FH models?
- C** One way of understanding it is by saying that it is the knowledge that the system has, but has not computed yet. More generally, it is knowledge that is logically implied by what the agent explicitly knows, and that could be reached after some inferential steps. In fact, implicit knowledge was originally introduced in computer science as part of a larger program aimed at addressing the logical omniscience problem.<sup>10</sup> In Kripke structures, knowledge is closed under logical implication. This places unrealistic computational demands on the reasoning abilities of agents. In economics, you talk about bounded rationality since Herbert Simon.<sup>11</sup> Similarly,

---

<sup>8</sup>See for instance, Mertens and Zamir (1984).

<sup>9</sup>Meier and Schipper (2014a).

<sup>10</sup>Levesque (1984) and Lakemeyer (1986). For logical omniscience problem, see also Hintikka (1975) and Stalnaker (1991). For more recent work, see Velázquez-Quesada (2013, 2014).

<sup>11</sup>Simon (1957).

in logic we are interested in logical non-omniscience and the distinction between explicit and implicit knowledge allows us to capture some it.

- E** That's not the notion of implicit knowledge in HMS models in which awareness is generated by primitive propositions. In those structures, both implicit and explicit knowledge are closed under logical implication, the difference is only the latter is closed only with respect to formulas involving primitive propositions of which the agent is aware. The limits to awareness are fully determined by the primitive propositions the agent is aware of and not by her procedural limits to conceive and reason about events. I do not really see how it is related to Herbert Simon's notions of bounded or procedural rationality, who very much emphasized the limits to the cognitive *processes*.<sup>12</sup>
- C** That's correct in so far we limit ourselves to awareness structures featuring awareness generated by primitive propositions. However, awareness structures offer more generally a flexible model in which limits to reasoning processes could be explicitly captured.<sup>13</sup>
- E** Awareness generated by primitive propositions is the notion of awareness mostly used in applications so far. With such a notion, unawareness is not a form of bounded rationality but rather bounded perception. With respect to what the agent is aware of, she is fully rational when it comes to information processing. So let me rephrase my question: What really is implicit knowledge in FH models with awareness generated by primitive propositions?
- C** Implicit knowledge that is not explicit knowledge is knowledge that the system would have if it were aware of it.
- E** I don't know what to make of such counterfactual knowledge. If the system were indeed aware of it, then its "state of mind" would be obviously different and with it perhaps also its knowledge. What necessitates that the system upon becoming aware would have explicit knowledge that is the same as its pre-aware implicit knowledge?
- C** Think of it more as comparative statics exercise that you often do in economics. There are two comparable "states of mind" with the same implicit knowledge. They just differ by awareness and hence the explicit knowledge. By studying how the agent's behavior changes when changing *only* awareness, you can isolate the causal effect of awareness keeping knowledge fixed.
- E** That would be a very useful exercise indeed. However, FH models do not necessarily model such comparable states of mind. Given a state in an awareness structure, there is not necessarily another state with the same knowledge "modulo awareness". This gives me an idea though: In HMS models, we model by construction comparable states of mind that just differ by awareness via projections

---

<sup>12</sup>Selten (2002).

<sup>13</sup>See Fagin and Halpern (1988), pp. 54-55, and Fagin et al. (1995), Chapter 9.5.

of states across the lattice of spaces. We can use them to make the comparative statics precise.

- C** This would be interesting. Yet, I do not see how you can accomplish this without a notion of implicit knowledge in HMS models.
- E** It should be possible to complement HMS models with a notion of implicit knowledge. However, in contrast to FH models, the notion of implicit knowledge could be a derived notion rather than the primitive notion.
- C** You could define “implicit” possibility sets by the inverse image of possibility sets in richer spaces in HMS models. You need to make sure though that such a construction yields a notion of implicit knowledge that preserves properties of “explicit” possibility sets such as transitivity and Euclideaness.
- E** Indeed. Figuring out the precise conditions under which we can derive implicit knowledge from explicit knowledge is non-trivial. However, it could help us to better understand the notion of implicit knowledge. For instance, a priori it is not clear that there is a unique notion of implicit knowledge that is consistent with explicit knowledge. And this leads me back to the question of whether it is even possible to test different notions of implicit knowledge and empirically identify them in data.
- C** I am not sure whether this is necessarily a meaningful question to ask in the context of knowledge of distributed systems. From the construction of the system, the system designer may have a pretty good idea about the way in which implicit knowledge is consistent with the system’s explicit knowledge. He designed the system and with it the notion of implicit knowledge. Moreover, the existence of behavioral implications of a notion of knowledge should just be *one* criterion for judging its relevance and properties. If a notion helps us to generate insights, we should study it irrespective of whether or not it has direct behavioral implications. Unobservable primitives have turned out to be useful in economics as well. Just look at utility theory. There you start with unobservable utilities or preferences as primitives.
- E** That’s why we are interested in revealed preferences. What would be the theory of “revealed implicit knowledge”? How can I conclude from observing a system that it has this or that implicit knowledge?
- C** Since implicit knowledge is knowledge that the system would have if it were more aware, you can raise the system’s awareness of some proposition and test it’s knowledge about it...
- E** ...but then I would just test explicit knowledge. Such a test procedure would destroy the very knowledge that I like to test for, namely implicit knowledge that is not explicit yet. That’s somewhat reminiscent of Heisenberg’s uncertainty principle in quantum physics, where the measurement destroys what is measured.



- C** We do not need to look as far as quantum physics. For instance, even in unawareness structures, it should be nearly impossible to test what the agent is *unaware* of without raising her awareness of the very proposition you want to test for and thereby destroying her unawareness. So already in unawareness structures used in economic applications, unawareness is not testable.
- E** That’s not the case. We should not assume that whatever an agent chooses must automatically raise her awareness of the contingencies on which the consequences of her choice depend on. In real life, we often make choices whose consequences we cannot fully anticipate. In a choice experiment, by letting an agent choose among contracts that only vary the consequences for a contingency and the negation of that contingency whose awareness I like to test for, I can observe whether the contingency or its negation becomes relevant to the agent’s choice. If neither the contingency nor its negation affect choices, then I can conclude from her choices that she must be unaware of the contingency.<sup>14</sup>
- C** I see, but it also seems to rest on the assumption that implicit knowledge could not affect choices ...
- P** Pardon me if I intrude on your conversation, but I overheard you talking about behavioral implications of implicit knowledge. The idea that the unconscious can affect behavior goes back at least to Freud in psychoanalysis (Freud, 1915). In modern psychology, there was a renaissance of the idea in the 1980s in the literature on implicit cognition, priming, and implicit measures.<sup>15</sup> You might be familiar with it if you took previously an implicit association test.<sup>16</sup> These tests typically reveal some discrimination based on some implicit beliefs of which the individual may not be aware of herself ...
- E** ... Isn’t this literature on implicit measures part of the replication crisis in psychology?
- P** Indeed, this literature has received a fair share of criticism both in academia and the media. For instance, famous priming studies failed to replicate.<sup>17</sup> Statistical irregularities were discovered.<sup>18</sup> Reported correlations on the implicit association test decreased over time.<sup>19</sup> Nevertheless, also statistically small effects can have large effects in society.<sup>20</sup> Moreover, given that there is quite some measurement error in such implicit measures, we shouldn’t expect to find large significant effects

---

<sup>14</sup>See Schipper (2013) for such a test.

<sup>15</sup>See Brownstein et al. (2019) and Gawronski and Hahn (2019).

<sup>16</sup>Greenwald et al. (1998).

<sup>17</sup>For instance, Doyen et al. (2012) failed to replicate the famous study by Bargh et al. (1996) in which participants unconsciously primed with age walked slower out of the lab.

<sup>18</sup>See for instance, Schimmack, Heene, and Kesavan (2017)

<sup>19</sup>Schimmack (2021) reevaluates the predictive validity of the Race Implicit Association Test and notes that the reported correlations declined over time from  $r = 0.38$  to  $r = 0.097$ . See also Meissner et al. (2019).

<sup>20</sup>This has been argued with regard to the Implicit Association Test by Greenwald, Banaji, and Nosek (2015).

even if they are there because of the attenuation bias. Finally, there is even disagreement on the very notion of implicit measures, whether they should be understood as measuring automatic biases<sup>21</sup> or unconscious biases<sup>22</sup>. So, I believe more careful research is needed in order to get a better grasp of implicit measures.

- E**    Hmm... Well, at least the literature offers an idea of how implicit information could have direct behavioral effects without the detour via explicit knowledge. I guess my earlier statement that only the explicit can affect behavior could be premature.
- C**    You mentioned the implicit association test and the implicit biases it might reveal. I was wondering whether an implicit bias doesn't point to mistakes in information processing, as people often disagree completely with their revealed biases. If that's the case, then probably an empirically valid notion of implicit belief should satisfy less stringent properties than explicit belief.
- E**    That's quite intuitive: If implicit knowledge refers to knowledge that is not necessarily present in the agent's mind and that she could not have revised or discussed with others, then this knowledge may involve mistakes in information processing as captured by violations of transitivity or Euclideaness of the "implicit" possibility sets. In contrast, explicit knowledge is the result of the agent's own reasoning and of discussions she could have had with others. Thus, we should rather expect explicit knowledge to satisfy strong properties such as S5.<sup>23</sup> Explicit knowledge is less prone to mistakes than implicit knowledge simply because the latter lies beneath some level of consciousness so that the individual could not have thought about it carefully.
- N**    Well, this runs counter to the received view in neuroscience. The brain is essentially a machine that minimizes consciousness.<sup>24</sup> Processes are as much as possible automated. You do not need to explicitly know everything involved when walking, even in difficult terrain. But you need to consciously reason for instance about a math problem you are trying to solve while walking home! The brain just requires explicit conscious reasoning for "non-standard" situations.<sup>25</sup> This would suggest that explicit knowledge is prone to mistakes while implicit knowledge is already sophisticated.
- E**    This is another argument for why explicit knowledge and implicit knowledge should not just differ by awareness. The implicit and the explicit "minds" seem to have

---

<sup>21</sup>Schimmack (2022)

<sup>22</sup>Gawronski, Ledgerwood, and Eastwick (2022). See also the commentaries on that article in *Psychological Inquiry*.

<sup>23</sup>"Strong S5 properties" refer to reflexivity and Euclideaness, corresponding to factivity and positive and negative introspection properties classically attributed to knowledge; see Rendsvig and Symons (2021).

<sup>24</sup>Roth (2003, pp. 236-240)

<sup>25</sup>*Ibid.*

different reasoning powers associated to them depending on how we interpret these notions of knowledge.

- P** Automated implicit knowledge reminds me of Gigerenzer’s notion of ecological rationality.<sup>26</sup> Like the ecological rationality of the individual, knowledge may match to the environment frequently encountered by the individual without such knowledge necessarily entering consciousness. But does it necessarily imply that it is sophisticated? If I am not wrong, mistakes in information processing are ruled out by the introspection axioms in the axiomatic system S5. However, these are precisely the axioms that the subject knows that she doesn’t know and that she knows that she does know. Doesn’t it imply self-consciousness of knowledge? I don’t see how it could possibly hold for ecological knowledge or implicit knowledge. Just consider the priming and implicit memory literature<sup>27</sup>, where subjects are shown quickly some information that they then show to have but not recall having. This clearly violates that subjects know that they know.
- C** Not necessarily. This just shows that they do not *explicitly* know what they know implicitly. But they could still implicitly know that they know it implicitly.
- P** Hmm, this reminds me of studies that showed that people were able to predict their own implicit bias pretty well.<sup>28</sup> This would suggest that in some sense they are aware of their implicit bias and that its not an unconscious phenomenon.
- S** To me it sounds like implicit knowledge could be interpreted as tacit knowledge à la Michael Polanyi<sup>29</sup> Tacit knowledge is knowledge that the individual has, but she doesn’t have the words to express. In this sense it is implicit.
- C** That sounds like an interesting interpretation of implicit knowledge.
- E** How could you test for what the individual knows only tacitly?
- S** Tacit knowledge affects behavior. An opera star obviously knows how to sing but she may not necessarily be able to describe how to sing. Similarly, Stradivari knew how to build excellent violins but was apparently unable to explicitly teach it to his successors.
- N** I cannot help noticing that in these examples the implicit tacit knowledge is the more sophisticated notion rather than the explicable knowledge. That conforms very much to how we think about implicit routine knowledge in neuroscience.
- E** To me the interpretation of tacit knowledge differs from implicit knowledge. Explicit knowledge is distinguished from implicit knowledge by awareness (at least in the formal literature on awareness), while “explicated” knowledge is apparently

---

<sup>26</sup>E.g., Gigerenzer (2002).

<sup>27</sup>Roediger and McDermott (1993).

<sup>28</sup>Hahn et al. (2014).

<sup>29</sup>Polanyi (1962).

distinguished from tacit knowledge by the ability to explicate or communicate it, not by awareness.

- C** I wonder whether at a formal level, a theory of “explicability” would look very different from a theory of awareness. After all, awareness generated by primitive propositions is defined by the set of primitive propositions on which the agent’s language is built. So implicit knowledge may also be interpretable as knowledge that the agent has but doesn’t have the language to express, i.e., tacit knowledge.
- E** Well, all those notions of implicit knowledge seem to have in common that they are beneath (or better beyond) some level of awareness or explicability. Perhaps HMS models are a useful tool to study them because the lattice structure reflects levels of awareness/explicability. By fitting them into decision and game theoretic models, we could develop rigorous testable theories of tacit knowledge and of implicit biases for instance. This would be a relevant enterprise.
- C** To implicit knowledge!

**All** Cheers!

The dialogue illustrates differences between FH and HMS models and more generally differences in the “modeling cultures” of computer science and economics. At the same time, it demonstrates that the notions of knowledge that arise in these structures can be interesting and relevant to various disciplines. Finally, it outlines an agenda of studying implicit knowledge. We take a first step in this paper. While the following sections focus on introducing implicit knowledge à la Fagin and Halpern (1988) in HMS models, we are more broadly interested on how to understand the notion of implicit knowledge itself. In the current paper, we aim for a notion of implicit knowledge that satisfies the strong properties (i.e., S5) and leave generalizations to future work.

In this paper, we show how to define implicit knowledge in HMS models so that it is consistent explicit knowledge, thereby providing a way to understand implicit knowledge in terms of explicit knowledge. We are aware of only a few other approaches deriving implicit knowledge from explicit knowledge. Using neighborhood models without a notion of awareness, Velázquez-Quesada (2013) takes explicit knowledge as the primitive and then derives implicit knowledge as closure of logical consequences of explicit knowledge. Implicit knowledge is then understood as knowledge that the agent ideally could deduce from her explicit knowledge. Lorini (2020) takes an agent’s belief base as explicit knowledge and derives implicit knowledge as what is deducible from an agent’s belief base and common background information. While we find these two notions of implicit knowledge easy to interpret, it is not the notion of implicit knowledge that is captured by FH models when awareness is generated by a set of primitive propositions, as argued by the economist in the dialogue. We also introduce a variant of HMS models in which we take the notion of implicit knowledge and a semantic awareness function as the primitive, and then derive explicit knowledge (we call these *implicit-knowledge based HMS models*). This shows that in HMS models, implicit and explicit knowledge are “interdefinable”, at

least in the sense that taking any of the two as primitive allows us to consistently specify the other, so one may choose either one as a primitive.

We are also interested in an extension of FH models that allows us to interpret them as subjective views of agents. Starting from an FH model, we show how to obtain such subjective views by forming a category of FH models with FH models as objects and surjective bounded morphisms as morphisms. Each category of FH models is literally a category of FH models that are modally equivalent relative to sublanguages formed by taking subsets of atomic formulas. The category of FH models forms a complete lattice ordered by subset inclusion on sets of atomic formulas or ordered by surjective bounded morphisms or ordered by modal equivalence relative to sublanguages. Each FH model in the lattice can be interpreted as the subjective model of an agent where her awareness level is given by the subset of atomic formulas over which the FH model is defined. The construction now suggests transformations between FH and HMS models. The transformation from FH to HMS models relies on a transformation of each FH category into an implicit knowledge-based HMS model mentioned above. This implicit knowledge-based HMS model can be complemented with explicit knowledge and thus yields a HMS model. The transformation from HMS to FH models simply relies on pruning away the subjective spaces, only maintaining the upmost space in the lattice, as well as deriving the syntactic awareness correspondences from possibility correspondences and the lattice of spaces in HMS models. For each model class, its transformation into a model of the other class satisfies the same formulas from a language for explicit, implicit knowledge, and awareness. As a corollary of soundness and completeness of the Logic of Propositional Awareness with respect to the class of FH models, the results allow us to derive soundness and completeness for the class of HMS models *now with implicit knowledge*, complementing earlier axiomatizations of HMS models that made use of explicit knowledge (and awareness) only (Heifetz, Meier, and Schipper, 2008, Halpern and Rêgo, 2008).

## 2 HMS Models

HMS models are multi-agent models for awareness originally proposed by Heifetz, Meier, and Schipper (2006, 2008). The goal was to provide a syntax-free semantics of multi-person unawareness that would fit seamlessly in decision and game theory. The model also served as an answer to the so-called impossibility result by Dekel, Lipman, and Rustichini (1998), who showed that state space models cannot capture non-trivial unawareness. HMS models recalled in this section escape the triviality by considering a lattice of spaces while FH models discussed in later sections do so by adding a syntactic awareness correspondence for each agent.

Throughout the paper, we let  $\text{At}$  be a non-empty set of atomic formulas.

**Definition 1 (HMS Model)** A HMS model  $\mathbf{M} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Pi_i)_{i \in I}, v \rangle$  for  $\text{At}$  consists of

- a non-empty set of individuals  $I$ ,

- a non-empty collection of non-empty disjoint state spaces  $\{S_\Phi\}_{\Phi \subseteq \text{At}}$  indexed by subsets of atomic formulas  $\Phi \subseteq \text{At}$ . Note that  $\{S_\Phi\}_{\Phi \subseteq \text{At}}$  forms a complete lattice by subset inclusion on the set of atomic formulas  $\Phi \subseteq \text{At}$ . Denote the set of all states in spaces of the lattice by  $\Omega := \bigcup_{\Phi \subseteq \text{At}} S_\Phi$ .
- Projections  $(r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}$  such that, for any  $\Phi, \Psi \subseteq \text{At}$  with  $\Psi \subseteq \Phi$ ,  $r_\Psi^\Phi : S_\Phi \rightarrow S_\Psi$  is surjective, for any  $\Phi \subseteq \text{At}$ ,  $r_\Phi^\Phi$  is the identity on  $S_\Phi$ , and for any  $\Phi, \Psi, \Upsilon \subseteq \text{At}$ ,  $\Upsilon \subseteq \Psi \subseteq \Phi$ ,  $r_\Upsilon^\Phi = r_\Upsilon^\Psi \circ r_\Psi^\Phi$ . For any  $\Phi \subseteq \text{At}$  and  $D \subseteq S_\Phi$ , denote by  $D^\uparrow := \bigcup_{\Phi \subseteq \Psi \subseteq \text{At}} (r_\Phi^\Psi)^{-1}(D)$ . An event  $E \subseteq \Omega$  is defined by a subset  $\Phi \subseteq \text{At}$  and a subset  $D \subseteq S_\Phi$  such that  $E := D^\uparrow$ . We call  $S_\Phi$  the base-space of the event  $E$  and  $D$  the base of the event  $E$ . We denote by  $\Sigma$  the set of events.
- A possibility correspondence  $\Pi_i : \Omega \rightarrow 2^\Omega \setminus \{\emptyset\}$  for each individual  $i \in I$ .
- A valuation function  $v : \text{At} \rightarrow \Sigma$ .

While Heifetz, Meier, and Schipper (2006) allowed for any non-empty complete lattice of non-empty disjoint spaces, we consider a lattice in which spaces are indexed by subsets of atomic formulas  $\Phi \subseteq \text{At}$  and the lattice order is induced by subset inclusion on the set of atomic formulas. This anticipates the connection to syntax (introduced by Heifetz, Meier, and Schipper, 2008) used in later sections. Implicitly, we can consider states in a space  $S_\Phi$  as consistent subsets of formulas in a language formed from the subset of atomic formulas in  $\Phi \subseteq \text{At}$ . Higher spaces in the lattice are associated with richer sets of atomic formulas and so richer sublanguages. The partial order between the state spaces induced by subset inclusion is thus also an order between expressiveness of languages. We refer for further discussions to Heifetz, Meier, and Schipper (2006, 2008).

Not every subset of the union of spaces is an event. Intuitively,  $D^\uparrow$  collects all the “extensions of descriptions in  $D$  to at least as expressive vocabularies” (Heifetz, Meier, and Schipper, 2006). Events are well defined by the above definition except for the case of vacuous events. Since the empty set is a subset of any space, we have as many vacuous events as there are state spaces. These vacuous events are distinguished by their base-space, so we denote them by  $\emptyset^{S_\Phi}$  for  $\Phi \subseteq \text{At}$ . At a first glance, the existence of many vacuous events may seem puzzling. Note that vacuous events essentially represent contradictions, i.e., propositions that cannot hold at any state. Contradictions are formed with atomic formulas. Thus, they can be more or less complicated depending on the expressiveness of the underlying language describing states and hence are represented by different vacuous events.

We define Boolean operations on events. Negation of events is defined as follows: Let  $E$  be an event with base  $D$  and base-space  $S_\Phi$ . Then  $\neg E := (S_\Phi \setminus D)^\uparrow$ . Conjunction of events is defined by intersection of events. Disjunction of events is defined by the DeMorgan Law using negation and conjunction as just defined. Note that in HMS models we typically have  $E \cup \neg E \subsetneq \Omega$  unless the base-space of  $E$  is  $S_\emptyset$ , the meet of the lattice of spaces. Also, disjunction of two events is typically a proper subset of the union of these events unless both events have the same base-space, since it is just the union of the events in spaces in which both events are “expressible”.

The following notation will be convenient: Sometimes we denote by  $S_\omega$  the state space that contains state  $\omega$ . For any  $D \subseteq S_\Phi$ , we denote by  $D_{S_\Psi}$  the projection of  $D$  to  $S_\Psi$  for  $\Psi \subseteq \Phi \subseteq \text{At}$ . We simplify notation further and let for any  $D \subseteq S_\Phi$  and  $\Psi \subseteq \Phi \subseteq \text{At}$ ,  $D_\Psi$  be the projection of  $D$  to  $S_\Psi$ . Similarly, for any  $D \subseteq S_\Psi$ , we denote by  $D^\Phi$  the “elaboration” of  $D$  in the space  $S_\Phi$  with  $\Psi \subseteq \Phi$ , i.e.,  $D^\Phi := (r_\Psi^\Phi)^{-1}(D)$ . The same applies to states, i.e.,  $\omega_\Psi$  is the projection of  $\omega \in S_\Phi$  to  $S_\Psi$  with  $\Psi \subseteq \Phi$ . Finally, for any event  $E \in \Sigma$ , we denote by  $S(E)$  the base-space of  $E$ . We say that an event  $E$  is expressible in  $S_\Phi$  if  $S(E) \preceq S_\Phi$ .

As usual in epistemic structures used in game theory and economics, information is modeled by a possibility correspondence instead of an accessibility relation. In HMS models, having mappings rather than relations adds extra convenience in that we can easily compose projections with possibility correspondences and vice versa. It is precisely the projective structure that makes HMS models tractable in applications and lets us analyze phenomena across “awareness levels”  $\{S_\Phi\}_{\Phi \subseteq \text{At}}$ . Since the motivation for HMS models in game theory and economics is to isolate the behavioral implications of unawareness from other factors like mistakes in information processing etc., we require that possibility correspondences satisfy strong properties analogous to S5.<sup>30</sup>

**Assumption 1 (Properties of the Possibility Correspondence)** *For any individual  $i \in I$ , we require that the possibility correspondence  $\Pi_i$  satisfies*

Confinement: *If  $\omega \in S_\Phi$ , then  $\Pi_i(\omega) \subseteq S_\Psi$  for some  $\Psi \subseteq \Phi$ .*

Generalized Reflexivity:  *$\omega \in \Pi_i^\uparrow(\omega)$  for every  $\omega \in \Omega$ .*<sup>31</sup>

Stationarity:  *$\omega' \in \Pi_i(\omega)$  implies  $\Pi_i(\omega') = \Pi_i(\omega)$ .*

Projections Preserve Ignorance: *If  $\omega \in S_\Phi$  and  $\Psi \subseteq \Phi$ , then  $\Pi_i^\uparrow(\omega) \subseteq \Pi_i^\uparrow(\omega_\Psi)$ .*

Projections Preserve Knowledge: *If  $\Upsilon \subseteq \Psi \subseteq \Phi$ ,  $\omega \in S_\Phi$  and  $\Pi_i(\omega) \subseteq S_\Psi$  then  $(\Pi_i(\omega))_\Upsilon = \Pi_i(\omega_\Upsilon)$ .*

Sometimes we denote by  $S_{\Pi_i(\omega)}$  the state space  $S$  for which  $\Pi_i(\omega) \subseteq S$ .

Stationarity corresponds to both transitivity and Euclideaness, while together Generalized Reflexivity and Stationarity corresponds to partitional properties in Kripke frames or Aumann structures. Note though that  $\Pi_i$  does not necessarily form a partition of a state space because we allow  $S_{\Pi_i(\omega)} \preceq S_\omega$ . We refer to Heifetz, Meier, and Schipper (2006, 2008) for further discussions of these properties.

Given the possibility correspondence, the knowledge operator is essentially defined like in Aumann (1999).

<sup>30</sup>Again, we refer to Heifetz, Meier, and Schipper (2006, 2008) for discussions of these properties. Generalizations are considered by Heifetz, Meier, and Schipper (2013a), Halpern and Rêgo (2008), Board, Chung, and Schipper (2011), and Galanis (2011, 2013).

<sup>31</sup>Here and in what follows, we abuse notation slightly and write  $\Pi_i^\uparrow(\omega)$  for  $(\Pi_i(\omega))^\uparrow$ .

**Definition 2 (Knowledge Operator)** For every individual  $i \in I$ , the knowledge operator on events is defined by, for every event  $E \in \Sigma$ ,

$$K_i(E) := \{\omega \in \Omega : \Pi_i(\omega) \subseteq E\}$$

if there exists a state  $\omega \in \Omega$  such that  $\Pi_i(\omega) \subseteq E$ , and by  $K_i(E) = \emptyset^{S(E)}$  otherwise.

**Definition 3 (Awareness Operator)** For every individual  $i \in I$ , the awareness operator on events is defined by, for every event  $E \in \Sigma$ ,

$$A_i(E) := \{\omega \in \Omega : S_{\Pi_i(\omega)} \succeq S(E)\}$$

if there exists a state  $\omega \in \Omega$  such that  $S_{\Pi_i(\omega)} \succeq S(E)$ , and by  $A_i(E) = \emptyset^{S(E)}$  otherwise. The unawareness operator is defined by  $U_i(E) := \neg A_i(E)$ .

We read  $K_i(E)$  as “individual  $i$  knows the event  $E$ ” and  $A_i(E)$  as “individual  $i$  is aware of event  $E$ ”.

**Lemma 1 (Heifetz, Meier, and Schipper, 2006)** For every individual  $i \in I$  and event  $E \in \Sigma$ , both  $K_i(E)$  and  $A_i(E)$  are  $S(E)$ -based events.

**Proposition 1 (Heifetz, Meier, and Schipper, 2006)** For every individual  $i \in I$ , the knowledge operator  $K_i$  satisfies the following properties: For every  $E, F \in \Sigma$  and  $\{E_n\}_n \subseteq \Sigma$ ,

- (i) Necessitation:  $K_i(\Omega) = \Omega$ ,
- (ii) Conjunction:  $K_i(\bigcap_n E_n) = \bigcap_n K_i(E_n)$ ,
- (iii) Truth:  $K_i(E) \subseteq E$ ,
- (iv) Positive Introspection:  $K_i(E) \subseteq K_i K_i(E)$ ,
- (v) Monotonicity:  $E \subseteq F$  implies  $K_i(E) \subseteq K_i(F)$ .
- (vi) Weak Negative Introspection I:  $\neg K_i(E) \cap \neg K_i \neg K_i(E) \subseteq \neg K_i \neg K_i \neg K_i(E)$ .

**Proposition 2 (Heifetz, Meier, and Schipper, 2006)** For every individual  $i \in I$ , the following properties of knowledge and awareness obtain: For every  $E \in \Sigma$  and  $\{E_n\}_n \subseteq \Sigma$ ,

1. KU Introspection:  $K_i U_i(E) = \emptyset^{S(E)}$ ,
2. AU Introspection:  $U_i(E) = U_i U_i(E)$
3. Weak Necessitation:  $A_i(E) = K_i(S(E)^\uparrow)$ ,
4. Plausibility:  $A_i(E) = K_i(E) \cup K_i \neg K_i(E)$ ,



5. *Strong Plausibility*:  $U_i(E) = \bigcap_{n=1}^{\infty} (\neg K_i)^n(E)$ ,
6. *Weak Negative Introspection II*:  $\neg K_i(E) \cap A_i \neg K_i(E) = K_i \neg K_i(E)$ ,
7. *Symmetry*:  $A_i(E) = A_i(\neg E)$ ,
8. *A-Conjunction*:  $\bigcap_n A_i(E_n) = A_i(\bigcap_n E_n)$ ,
9. *AK-Self Reflection*:  $A_i(E) = A_i K_i(E)$ ,
10. *AA-Self Reflection*:  $A_i(E) = A_i A_i(E)$ ,
11. *A-Introspection*:  $A_i(E) = K_i A_i(E)$ .

Heifetz, Meier, and Schipper (2006, 2008) also define mutual knowledge, common knowledge, mutual awareness, and common awareness operators and derive properties, which we skip here for brevity.

The following lemma turns out to be very useful but has not been proved in the literature. It proves a consistency condition of the possibility correspondence. To see the content of the lemma, consider any state  $\omega$  and its projection (“less expressive version”)  $\omega_\Psi$ . If at  $\omega$  the individual’s knowledge and awareness is given by  $\Pi_i(\omega) \subseteq S_\Upsilon$ , for some  $S_\Upsilon$  that involves even less awareness than  $S_\Psi$  (i.e., where  $\Upsilon \subseteq \Psi$ ), then at  $\omega_\Psi$  the individual’s knowledge and awareness is also given by  $\Pi_i(\omega)$ . That is, at a less expressive projection  $\omega_\Psi$ , the individual cannot be aware of more than she is at  $\omega$  and cannot have knowledge that differs from her knowledge at  $\omega$ .

**Lemma 2** *For every individual  $i \in I$  and any  $\Upsilon \subseteq \Psi \subseteq \Phi \subseteq \text{At}$ , if  $\omega \in S_\Phi$  and  $\Pi_i(\omega) \subseteq S_\Upsilon$ , then  $\Pi_i(\omega_\Psi) = \Pi_i(\omega)$ .*

PROOF. By Projections Preserve Ignorance,  $\Pi_i^\uparrow(\omega_\Psi) \subseteq \Pi_i^\uparrow(\omega_\Upsilon)$  since  $(\omega_\Psi)_\Upsilon = \omega_\Upsilon$ . By Generalized Reflexivity,  $\omega_\Upsilon \in \Pi_i(\omega)$ . By Stationarity,  $\Pi_i(\omega_\Upsilon) = \Pi_i(\omega)$ . Thus,  $\Pi_i^\uparrow(\omega_\Psi) \subseteq \Pi_i^\uparrow(\omega)$ .

Also, if  $\Pi_i(\omega_\Psi) \subseteq S_\Upsilon$ , then  $\Pi_i(\omega_\Psi) = \Pi_i(\omega)$  follows from Stationarity. It now follows that if  $\Pi_i(\omega_\Psi) \neq \Pi_i(\omega)$ , we must have  $\Pi_i(\omega_\Psi) \subseteq S_\Delta$  with  $\Upsilon \subsetneq \Delta \subseteq \Psi$ . Then there exists an event  $E \in \Sigma$  with  $S(E) = S_\Upsilon$  s.t.  $\omega_\Psi \in A_i(E)$  and  $\omega \in U_i(E)$ . Since  $U_i(E)$  is an  $S(E)$  based event by Lemma 1 and the definition of negation, we must have  $\omega_\Psi \in U_i(E)$ , a contradiction to  $\omega_\Psi \in A_i(E)$ .  $\square$

### 3 From Explicit to Implicit Knowledge

In this section, we define the “implicit” possibility correspondence  $\Lambda_i$  such that it is consistent with the “explicit” possibility correspondence  $\Pi_i$ . We then define implicit knowledge as based on  $\Lambda_i$  and show that it satisfies standard S5 properties as well as properties of Fagin and Halpern (1988) that are jointly satisfied by implicit knowledge, explicit knowledge, and awareness.

From now on, for any individual  $i \in I$ , we call  $\Pi_i$  the *explicit* possibility correspondence,  $\Pi_i(\omega)$  *explicit* possibility set at  $\omega$ , and  $K_i(E)$  the event that  $i$  *explicitly* knows  $E$ .

**Definition 4 (Implicit Possibility Correspondence)** *Given the explicit possibility correspondence  $\Pi_i$  of individual  $i \in I$ , let the implicit possibility correspondence  $\Lambda_i : \Omega \longrightarrow 2^\Omega$  satisfy*

Reflexivity: *For any  $\omega \in \Omega$ ,  $\omega \in \Lambda_i(\omega)$ .*

Stationarity:  *$\omega' \in \Lambda_i(\omega)$  implies  $\Lambda_i(\omega') = \Lambda_i(\omega)$ .*

Projections Preserve Implicit Knowledge: *For any  $\Phi \subseteq \text{At}$ , if  $\omega \in S_\Phi$ , then  $\Lambda_i(\omega)_\Psi = \Lambda_i(\omega_\Psi)$  for all  $\Psi \subseteq \Phi$ .*

Explicit Measurability:  *$\omega' \in \Lambda_i(\omega)$  implies  $\Pi_i(\omega') = \Pi_i(\omega)$ .*

Implicit Measurability:  *$\omega' \in \Pi_i(\omega)$  implies  $\Lambda_i(\omega') = \Lambda_i(\omega)_{S_{\Pi_i(\omega)}}$ .*

**Definition 5 (Complemented HMS Model)** *Given an HMS model  $M = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Pi_i)_{i \in I}, v \rangle$  and a collection of implicit possibility correspondences  $(\Lambda_i)_{i \in I}$  satisfying the above properties, we call  $\bar{M} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Pi_i)_{i \in I}, (\Lambda_i)_{i \in I}, v \rangle$  a complemented HMS model.*

A complemented HMS model is a HMS model complemented with an implicit possibility correspondence for each individual. In the following, we discuss and derive properties of the implicit possibility correspondence. It also demonstrates ways in which the implicit possibility correspondence is consistent with the explicit possibility correspondence.

Reflexivity and Stationarity are standard and imply that  $\{\Lambda_i(\omega)\}_{\omega \in S_\Phi}$  forms a partition of  $S_\Phi$  for every  $\Phi \subseteq \text{At}$ . It is straightforward to see that they also imply a strengthening of Confinement: The implicit possibility set at a state must be a subset of the space in which the state itself lies. That is, both the state and the implicit possibility set are described using the *same* language. More formally:

**Remark 1 (Strong Confinement)** *For any individual  $i \in I$ ,  $\Phi \subseteq \text{At}$ , and  $\omega \in S_\Phi$ ,  $\Lambda_i(\omega) \subseteq S_\Phi$ .*

Projections Preserve Implicit Knowledge is analogous to Projections Preserve Knowledge satisfied by  $\Pi_i$ . The absence of Projections Preserve (Implicit) Ignorance from the above list of properties imposed to the implicit possibility correspondence may look puzzling at a first glance. Yet, as we show below it is implied by Strong Confinement and Projections Preserve Implicit Knowledge.

**Lemma 3 (Projections Preserve Implicit Ignorance)** *For any individual  $i \in I$ , if  $\Lambda_i$  satisfies Strong Confinement and Projections Preserve Implicit Knowledge, then  $\Lambda_i$  satisfies Projections Preserve Implicit Ignorance. That is, for all  $\Phi \subseteq \text{At}$ , if  $\omega \in S_\Phi$ , then  $\Lambda_i^\uparrow(\omega) \subseteq \Lambda_i^\uparrow(\omega_\Psi)$  for all  $\Psi \subseteq \Phi$ .*

PROOF. If  $\omega \in S_\Phi$ , then  $\Lambda_i(\omega) \subseteq S_\Phi$  by Strong Confinement. By Projections Preserve Implicit Knowledge,  $(\Lambda_i(\omega)_\Psi)^\uparrow = \Lambda_i^\uparrow(\omega_\Psi)$ . Since  $\Psi \subseteq \Phi$ , it follows now that  $\Lambda_i^\uparrow(\omega) \subseteq \Lambda_i^\uparrow(\omega_\Psi)$ .  $\square$

Explicit Measurability says that explicit knowledge is measurable with respect to implicit knowledge. That is, the agent always implicitly knows her explicit knowledge. The converse, Implicit Measurability, is more subtle because of awareness. An individual may not explicitly know her implicit knowledge because she might be unaware of some events. However, the individual always explicitly knows her implicit knowledge modulo awareness. That is, she might implicitly know more at a higher awareness level than what she knows at her awareness level (like in the structure to the right in Figure 1) but at her awareness level, her implicit knowledge equals her explicit knowledge. The following lemma formalizes the last conclusion. The proof uses all properties of  $\Pi_i$  and  $\Lambda_i$  except Projections Preserve Knowledge of both  $\Lambda_i$  and  $\Pi_i$  and Projections Preserve Ignorance of  $\Pi_i$ .

**Lemma 4** *For any individual  $i \in I$ , if  $\omega' \in \Pi_i(\omega)$ , then  $\Lambda_i(\omega') = \Pi_i(\omega')$ .*

PROOF. Let  $\omega' \in \Pi_i(\omega)$ . This implies  $\Pi_i(\omega') = \Pi_i(\omega)$  by Stationarity of  $\Pi_i$  and  $\Lambda_i(\omega') = \Lambda_i(\omega)_{S_{\Pi_i(\omega)}}$  by Implicit Measurability. By Generalized Reflexivity of  $\Pi_i$ , we have  $\omega' \in \Pi_i(\omega')$  and by Reflexivity of  $\Lambda_i$ , we have  $\omega' \in \Lambda_i(\omega')$ . Thus,  $\Pi_i(\omega') \cap \Lambda_i(\omega') \neq \emptyset$ .

Suppose that there exists  $\omega'' \in \Pi_i(\omega')$  with  $\omega'' \notin \Lambda_i(\omega')$ . Then by Stationarity of  $\Lambda_i$ ,  $\Lambda_i(\omega'') \cap \Lambda_i(\omega') = \emptyset$ . But  $\omega'' \in \Pi_i(\omega')$  implies in this case by Implicit Measurability that  $\Lambda_i(\omega'') = \Lambda_i(\omega)$ , a contradiction.

Now suppose that there exists  $\omega'' \in \Lambda_i(\omega')$  with  $\omega'' \notin \Pi_i(\omega')$ . By Stationarity of  $\Pi_i$ ,  $\Pi_i(\omega'') \cap \Pi_i(\omega') = \emptyset$ . But  $\omega'' \in \Lambda_i(\omega')$  implies by Explicit Measurability that  $\Pi_i(\omega'') = \Pi_i(\omega')$ , a contradiction.  $\square$

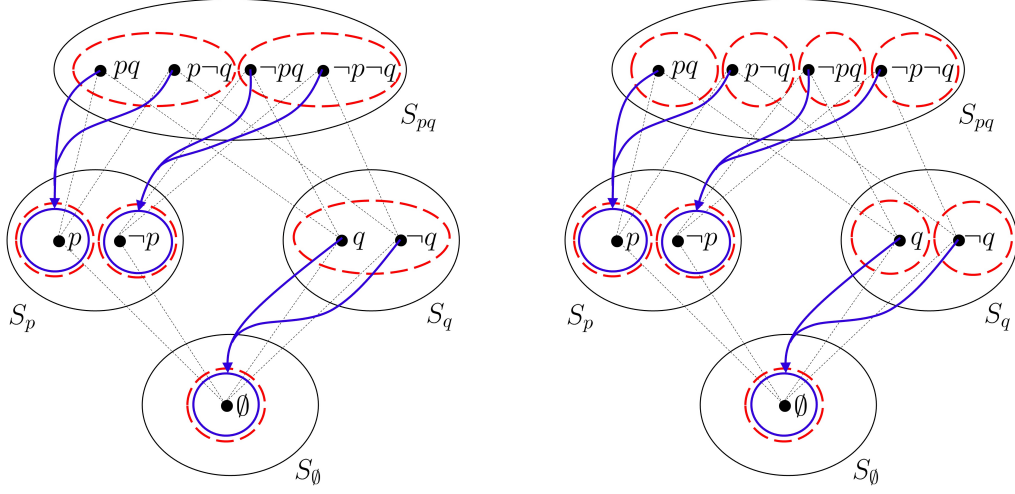
Above properties imply now a strong connection between implicit and explicit possibility sets:

**Lemma 5 (Coherence)** *For any individual  $i \in I$ ,  $\omega \in \Omega$ ,  $\Lambda_i(\omega)_{S_{\Pi_i(\omega)}} = \Pi_i(\omega)$ .*

PROOF. For all  $\omega \in \Omega$ , by Confinement and Generalized Reflexivity,  $\omega_{S_{\Pi_i(\omega)}} \in \Pi_i(\omega)$ . By Lemma 4,  $\Lambda_i(\omega_{S_{\Pi_i(\omega)}}) = \Pi_i(\omega_{S_{\Pi_i(\omega)}})$ . By Stationarity of  $\Pi_i$ ,  $\Pi_i(\omega_{S_{\Pi_i(\omega)}}) = \Pi_i(\omega)$ . Thus,  $\Lambda_i(\omega_{S_{\Pi_i(\omega)}}) = \Pi_i(\omega)$ . By Projections Preserve Implicit Knowledge of  $\Lambda_i$ ,  $\Lambda_i(\omega)_{S_{\Pi_i(\omega)}} = \Pi_i(\omega)$ .  $\square$

Figure 1 shows two single-agent examples of how implicit knowledge is fitted to explicit knowledge. Consider first the HMS model to the left. There are four spaces indexed by subsets of atomic formulas. Anticipating the semantics of HMS models introduced later, we describe and call states by their atomic formulas. The explicit possibility correspondence of the individual is indicated by the solid blue ovals and arrows. For

Figure 1: Two Single-Agent Examples of Complemented HMS Models



instance, at state  $pq$  she considers possible state  $p$ . That is, she is unaware of  $q$  and knows  $p$ . Similarly, at state  $\neg pq$  she is unaware of  $q$  and knows  $\neg p$ . Her implicit possibility correspondence is given by the red dashed ovals. Note that in this complemented HMS model she does not implicitly know more than she does explicitly. Contrast this with the HMS model to the right. There, she implicitly knows  $q$  for instance at state  $pq$  (because her implicit possibility set at  $pq$  is  $\{pq\}$ ) although she is not aware of  $q$  (because her explicit possibility set at  $pq$  is on  $S_{\{p\}}$ ). and hence does not explicitly know  $q$ . The figures demonstrate that the implicit possibility correspondence may be consistent with the explicit possibility correspondence in two different ways. It may model implicit knowledge that is finer than the explicit knowledge (like in the figure to the right) or implicit knowledge that is as coarse as the explicit knowledge but not coarser (like in the figure to the left). Note that a version of the models in Figure 1 in which only  $\{pq, p\neg q\}$  is in a red dashed oval while  $\neg pq$  and  $\neg p\neg q$  are in distinct red dashed circles in  $S_{pq}$  is ruled out by Projections Preserve Implicit Knowledge.

Given implicit possibility correspondences, we proceed with the definition of the implicit knowledge operators.

**Definition 6 (Implicit Knowledge Operator)** For any individual  $i \in I$ , the implicit knowledge operator on events  $E \in \Sigma$  is

$$L_i(E) := \{\omega \in \Omega : \Lambda_i(\omega) \subseteq E\}$$

if there exists a state  $\omega \in \Omega$  such that  $\Lambda_i(\omega) \subseteq E$  and by  $L_i(E) = \emptyset^{S(E)}$  otherwise.

The next observation follows immediately from the properties of the implicit possibility correspondence and the proof of Lemma 1.

**Lemma 6** For any individual  $i \in I$  and event  $E \in \Sigma$ ,  $L_i(E)$  is an  $S(E)$ -based event.

Implicit knowledge satisfies all properties of “partitional” knowledge.

**Proposition 3** *For any individual  $i \in I$ ,  $L_i$  satisfies for any  $E, F \in \Sigma$  and  $\{E_n\}_n \subseteq \Sigma$ ,*

- (i) *Necessitation: For  $\Phi \subseteq \mathbf{At}$ ,  $L_i(S_\Phi^\uparrow) = S_\Phi^\uparrow$ ,*
- (ii) *Conjunction:  $L_i(\bigcap_n E_n) = \bigcap_n L_i(E_n)$ ,*
- (iii) *Monotonicity:  $E \subseteq F$  implies  $L_i(E) \subseteq L_i(F)$ ,*
- (iv) *Truth:  $L_i(E) \subseteq E$ ,*
- (v) *Positive Introspection:  $L_i(E) \subseteq L_i L_i(E)$ ,*
- (vi) *Negative Introspection:  $\neg L_i(E) \subseteq L_i \neg L_i(E)$ .*

PROOF. All properties follow analogously from the proof of Proposition 1 in Heifetz, Meier, and Schipper (2006) except for Necessitation and Negative Introspection. Necessitation follows straightforwardly from Strong Confinement. Negative Introspection follows by standard arguments from Stationarity, e.g., Osborne and Rubinstein (1994), p. 70.  $\square$

We observe that, as in Fagin and Halpern (1988), explicit knowledge of an event equals to implicit knowledge and awareness of that event.

**Proposition 4** *For any individual  $i \in I$  and event  $E \in \Sigma$ ,*

- 1.  $K_i(E) = L_i(E) \cap A_i(E)$ ,
- 2.  $U_i(E) = L_i(U_i(E))$ ,
- 3.  $A_i(E) = L_i(A_i(E))$ ,
- 4.  $A_i L_i(E) = A_i(E)$ .

PROOF. 1. “ $\subseteq$ ”:  $\omega \in K_i(E)$  if and only if  $\Pi_i(\omega) \subseteq E$ . By Coherence (Lemma 5)  $\Lambda_i(\omega)_{S_{\Pi_i(\omega)}} = \Pi_i(\omega)$ . Thus,  $\Lambda_i(\omega) \subseteq E$ , which is equivalent to  $\omega \in L_i(E)$ . Moreover,  $K_i(E) \subseteq A_i(E)$ . Thus,  $\omega \in A_i(E)$ . Hence  $\omega \in L_i(E) \cap A_i(E)$ .

“ $\supseteq$ ”:  $\omega \in A_i(E)$  if and only if  $\Pi_i(\omega) \subseteq S \succeq S(E)$ . By Coherence (Lemma 5) of  $\Lambda_i$ ,  $\Lambda_i(\omega)_{S_{\Pi_i(\omega)}} \subseteq S \succeq S(E)$ .  $\omega \in L_i(E)$  if and only if  $\Lambda_i(\omega) \subseteq E$ . It now follows that  $\Lambda_i(\omega)_{S_{\Pi_i(\omega)}} \subseteq E$  and thus by Coherence (Lemma 5),  $\Pi_i(\omega) \subseteq E$ , which is equivalent to  $\omega \in K_i(E)$ .

2. “ $\subseteq$ ”: Let  $\omega \in U_i(E) = \neg A_i(E)$  and suppose by contradiction that  $\omega \notin L_i(U_i(E)) = L_i(\neg A_i(E))$ . This means that  $\Lambda_i(\omega) \not\subseteq \neg A(E)$ , i.e., that there exists some  $\omega' \in \Lambda_i(\omega)$  such that  $\omega' \notin \neg A(E)$ . Since  $\omega \in \neg A_i(E)$ , then  $S_{\Pi_i(\omega)} \not\subseteq S(E)$ . However, by Explicit Measurability, since  $\omega' \in \Lambda_i(\omega)$ , then  $\Pi_i(\omega) = \Pi_i(\omega')$ , which means that  $S_{\Pi_i(\omega')} \not\subseteq S(E)$ .

This amounts to saying that  $\omega' \in \neg A(E)$ , contradicting the assumption that  $\omega' \notin \neg A(E)$ . Hence,  $\omega \in L_i(U_i(E))$ .

“ $\supseteq$ ”: Immediate by Reflexivity of  $\Lambda_i$ .

3. “ $\subseteq$ ”: Let  $\omega \in A_i(E)$ . Then  $S_{\Pi_i(\omega)} \succeq S(E)$ . By Explicit Measurability, for all  $\omega' \in \Lambda_i(\omega)$ ,  $\Pi_i(\omega) = \Pi_i(\omega')$ , so for all  $\omega' \in \Lambda_i(\omega)$ ,  $S_{\Pi_i(\omega')} \succeq S(E)$ . This means that for all  $\omega' \in \Lambda_i(\omega)$ ,  $\omega' \in A_i(E)$ , i.e.,  $\Lambda_i(\omega) \subseteq A_i(E)$ . Thus,  $\omega \in L_i(A_i(E))$ .

“ $\supseteq$ ”: Immediate by Reflexivity of  $\Lambda_i$ .

4.  $A_i(L_i(E)) = K_i(S(L_i(E))^\uparrow) = K_i(S(E)^\uparrow) = A_i(E)$ , where the first and last equality follows from Weak Necessitation and the equality in the middle follows from Lemma 6.  $\square$

Properties 2. and 3. above mean that the individual implicitly knows her unawareness. This is an aspect where implicit knowledge differs from explicit knowledge. Indeed, by KU introspection (cf. Proposition 2), an individual can never explicitly know that she is unaware of an event. Property 4 says that an individual is aware of her implicit knowledge of an event if and only if she is aware of the event. That is, the moment she can reason about an event, she can also reason about her implicit knowledge of the event. This is analogous to AK-Self-Reflection of explicit knowledge.

## 4 From Implicit to Explicit Knowledge

In the previous section, we defined implicit knowledge based on explicit knowledge. In this section, we go the other direction. We can devise a version of HMS models that feature as primitives possibility correspondences capturing implicit knowledge and (non-syntactic) awareness functions, and then derive the possibility correspondence capturing explicit knowledge.

**Definition 7 (Implicit Knowledge-Based HMS Model)** *An implicit knowledge-based HMS model  $M^* = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  consists of*

- *a non-empty set of individuals  $I$ ,*
- *a nonempty collection of nonempty disjoint state spaces  $\{S_\Phi\}_{\Phi \subseteq \text{At}}$  (as in Definition 1),*
- *projections  $(r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}$  (as in Definition 1),*
- *an implicit possibility correspondence  $\Lambda_i^* : \Omega \longrightarrow 2^\Omega \setminus \{\emptyset\}$ , for all  $i \in I$ ,*
- *an awareness function  $\alpha_i : \Omega \longrightarrow \{S_\Phi\}_{\Phi \subseteq \text{At}}$ , for all  $i \in I$ ,*
- *a valuation function  $v : \text{At} \longrightarrow \Sigma$ .*

Like HMS models, implicit knowledge-based HMS models feature a projective lattice of state spaces. However, instead of the explicit possibility correspondence, we now take the implicit possibility correspondences as a primitive. As before, we are interested in strong properties of knowledge associated with S5 because (1) these properties have been used for explicit knowledge in applications, and (2) we will require explicit knowledge to be consistent with implicit knowledge. As such, we are interested in how the rich structure of S5 translates into properties of a derived explicit possibility correspondence. To that end, we require:

**Assumption 2 (Properties of Implicit Possibility Correspondences)** *For each individual  $i \in I$ , the implicit possibility correspondence  $\Lambda_i^*$  satisfies Reflexivity, Stationarity, and Projections Preserve Implicit Knowledge.*

These properties were also satisfied by implicit possibility correspondences in the previous section.<sup>32</sup>

The second primitive of implicit knowledge-based HMS models is the awareness function  $\alpha_i$  for every individual  $i \in I$ . We impose the following properties on  $\alpha_i$ :

**Assumption 3 (Properties of Awareness Functions)** *For each individual  $i \in I$ , the awareness function  $\alpha_i : \Omega \rightarrow \{S_\Phi\}_{\Phi \subseteq \text{At}}$  satisfies*

- O. Lack of Conception: If  $\omega \in S_\Phi$ , then  $\alpha_i(\omega) \preceq S_\Phi$ .*
- I. Awareness Measurability: If  $\omega' \in \Lambda_i^*(\omega)$ , then  $\alpha_i(\omega') = \alpha_i(\omega)$ .*
- II. If  $\omega \in S_\Phi$  and  $S_\Psi \preceq \alpha_i(\omega)$ , then  $\alpha_i(\omega_\Psi) = S_\Psi$ .*
- III. If  $\omega \in S_\Phi$  and  $\alpha_i(\omega) \preceq S_\Psi \preceq S_\Phi$ , then  $\alpha_i(\omega_\Psi) = \alpha_i(\omega)$ .*
- IV. If  $\omega \in S_\Phi$  and  $\Psi \subseteq \Phi$ , then  $\alpha_i(\omega) \succeq \alpha_i(\omega_\Psi)$ .*

When  $\alpha_i(\omega) \in S$  for some  $S \in \{S_\Phi\}_{\Phi \subseteq \text{At}}$ , we call  $S$  the awareness level of  $i$  at  $\omega$ .

Property O. models one feature of Confinement (cf. Assumption 1). Note that Confinement in HMS models has two features: First, it requires that the possibility set at a state is a subset of exactly one space. Second, it says that this space must be weakly less expressive than the space containing the state. Only this second feature is captured by property O. The idea is that an individual may have lack of conception. Property I. is a measurability condition. Awareness is measurable with respect to implicit knowledge. The implication is that an agent implicitly knows her own awareness. Properties II. to IV. are consistency properties of awareness across the lattice. Projections preserve awareness as long as it is still expressible in the spaces. While property II. preserves awareness

---

<sup>32</sup>Note again that Reflexivity and Stationarity implies Strong Confinement. In more general settings without Reflexivity or Stationarity, at least Strong Confinement would have to be imposed on  $\Lambda_i^*$  for every  $i \in I$ .

for corresponding states in spaces less expressive than the awareness level at a state, property III. preserves awareness for corresponding states in spaces more expressive than the awareness level at that state.

**Definition 8 (Derived Explicit Possibility Correspondence)** *Given an implicit knowledge-based HMS model  $M^* = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$ , define for each individual  $i \in I$  the explicit possibility correspondence  $\Pi_i^* : \Omega \longrightarrow 2^\Omega$  by, for all  $\omega \in \Omega$  and  $\Phi \subseteq \text{At}$ ,*

$$\Pi_i^*(\omega_\Phi) := \Lambda_i^*(\omega)_{\alpha_i(\omega_\Phi)}.$$

*We call  $\bar{M}^* = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  the complemented implicit knowledge-based HMS model.*

The defining condition for the explicit possibility correspondence in implicit knowledge-based HMS models is a slight strengthening of Coherence derived from the explicit and implicit measurability in Lemma 5. Here we take it as the primitive to connect explicit knowledge to implicit knowledge.

The following observations are immediate:

**Lemma 7** *For all  $\omega \in \Omega$ ,*

- A.  $\Pi_i^*(\omega) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$ ,
- B.  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_\Phi$  for all  $\Phi \subseteq \text{At}$  with  $S_\Phi \preceq \alpha_i(\omega)$ ,
- C.  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$  for all  $\Phi \subseteq \text{At}$  with  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$ .

PROOF. A. For any  $\Psi \subseteq \text{At}$ , if  $\omega \in S_\Psi$ , then  $\Pi_i^*(\omega) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$  by definition.

B. For all  $\Phi \subseteq \text{At}$ ,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_{\alpha_i(\omega_\Phi)}$  by definition. Since  $S_\Phi \preceq \alpha_i(\omega)$ , by II.  $\alpha_i(\omega_\Phi) = S_\Phi$ . Thus,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_\Phi$ .

C. Again, for all  $\Phi \subseteq \text{At}$ ,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_{\alpha_i(\omega_\Phi)}$  by definition. Since in this case  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$ , III. implies  $\alpha_i(\omega_\Phi) = \alpha_i(\omega)$ . Hence,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$ .  $\square$

The following sequence of lemmata records properties of the derived explicit possibility correspondence. It shows that it satisfies the properties of the explicit possibility correspondence of HMS models.

**Lemma 8** *For any individual  $i \in I$ , if  $\alpha_i$  satisfies O., II., and III., then  $\Pi_i^*$  satisfies Confinement.*

PROOF. If  $\omega \in \Omega$ , then by A.,  $\Pi_i^*(\omega) = \Lambda_i^*(\omega)_{\alpha_i(\omega)} \subseteq \alpha_i(\omega)$ . By O.,  $\alpha_i(\omega) \preceq S_\omega$ .

By C.,  $\Pi_i^*(\omega_\Phi) = \Pi_i^*(\omega)$  for  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$ .  $\Pi_i^*(\omega) \subseteq \alpha_i(\omega)$ .

By B.,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_\Phi$  for  $S_\Phi \preceq \alpha_i(\omega)$ . Hence  $\Pi_i^*(\omega_\Phi) \subseteq S_\Phi$ .  $\square$



**Lemma 9** *For any individual  $i \in I$ , if  $\alpha_i$  satisfies O., II., and III., then Reflexivity of  $\Lambda_i^*$  implies that  $\Pi_i^*$  satisfies Generalized Reflexivity.*

PROOF. By Reflexivity of  $\Lambda_i^*$ ,  $\omega \in \Lambda_i^*(\omega)$ . By O.,  $\Lambda_i^*(\omega) \subseteq (\Lambda_i^*(\omega)_{\alpha_i(\omega)})^\uparrow$ . Thus, by A.  $\omega \in (\Lambda_i^*(\omega)_{\alpha_i(\omega)})^\uparrow = \Pi_i^{*\uparrow}(\omega)$ . For any  $S_\Phi \preceq \alpha_i(\omega)$ ,  $\omega_\Phi \in \Lambda_i^*(\omega)_\Phi = \Pi_i^*(\omega_\Phi)$  by B. For any  $\Phi \subseteq \text{At}$  with  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$ ,  $\omega_\Phi \in (\Lambda_i^*(\omega)_{\alpha_i(\omega)})^\uparrow = \Pi_i^{*\uparrow}(\omega_\Phi)$  by C.  $\square$

**Lemma 10** *For any individual  $i \in I$ , if  $\alpha_i$  satisfies I., II., and III., then Stationarity of  $\Lambda_i^*$  implies that  $\Pi_i^*$  satisfies Stationarity.*

PROOF. For any  $\omega \in \Omega$ , A. implies  $\Lambda_i^*(\omega)_{\alpha_i(\omega)} = \Pi_i^*(\omega)$ . So we have  $\omega' \in \Pi_i^*(\omega)$  if and only if  $\omega' \in \Lambda_i^*(\omega)_{\alpha_i(\omega)}$ . Then we know that there exists  $\omega'' \in (r_{\alpha_i(\omega)}^{S_\omega})^{-1}(\{\omega'\})$  with  $\omega'' \in \Lambda_i^*(\omega)$ . By Stationarity of  $\Lambda_i^*$ ,  $\Lambda_i^*(\omega'') = \Lambda_i^*(\omega)$ . Thus, by I.  $\Lambda_i^*(\omega'')_{\alpha_i(\omega'')} = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$  and hence by A.,  $\Pi_i^*(\omega'') = \Pi_i^*(\omega)$ . Since  $\omega''_{\alpha_i(\omega'')} = \omega'$ , by B.  $\Pi_i^*(\omega') = \Pi_i^*(\omega)$ .

Now consider the case  $S_\Phi \preceq \alpha_i(\omega)$  and  $\omega' \in \Pi_i^*(\omega_\Phi)$ . The arguments are analogous to previous arguments except that from I. we conclude  $\Lambda_i^*(\omega'')_\Phi = \Lambda_i^*(\omega)_\Phi$ .

Finally, consider the case  $\Phi \subseteq \text{At}$  with  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$  and  $\omega' \in \Pi_i^*(\omega_\Phi)$ . The arguments are analogous to the arguments in the first paragraph of the proof except that  $\omega$  is replaced by  $\omega_\Phi$ .  $\square$

**Lemma 11** *For any individual  $i \in I$ , if  $\alpha_i$  satisfies II., III., and IV., then  $\Pi_i^*$  satisfies Projections Preserve Ignorance.*

PROOF. For any  $\omega \in \Omega$ , A. implies  $\Lambda_i^{*\uparrow}(\omega)_{\alpha_i(\omega)} = \Pi_i^{*\uparrow}(\omega)$ .

If  $S_\Psi \preceq \alpha_i(\omega)$ , then B. implies  $\Pi_i^{*\uparrow}(\omega) \subseteq \Lambda_i^{*\uparrow}(\omega)_\Psi = \Pi_i^{*\uparrow}(\omega_\Psi)$ .

If  $\Psi \subseteq \text{At}$  with  $S_\omega \succeq S_\Psi \succeq \alpha_i(\omega)$ , then C. implies  $\Pi_i^{*\uparrow}(\omega) = \Pi_i^{*\uparrow}(\omega_\Psi)$ .

If  $\Psi \subseteq \text{At}$  such that  $S_\omega \succeq S_\Psi$  and  $S_\Psi$  is incomparable to  $\alpha_i(\omega)$ , then  $\Lambda_i^{*\uparrow}(\omega)_{\alpha_i(\omega)} \subseteq \Lambda_i^{*\uparrow}(\omega)_{\alpha_i(\omega_\Psi)}$  since by IV.,  $\alpha_i(\omega) \succeq \alpha_i(\omega_\Psi)$ . Thus,  $\Pi_i^{*\uparrow}(\omega) \subseteq \Pi_i^{*\uparrow}(\omega_\Psi)$ .  $\square$

**Lemma 12** *For any individual  $i \in I$ , if  $\alpha_i$  satisfies II. and III., then  $\Pi_i^*$  satisfies Projections Preserve Knowledge.*

PROOF. For any  $\omega \in \Omega$ ,  $\Lambda_i^*(\omega)_{\alpha_i(\omega)} = \Pi_i^*(\omega) \subseteq \alpha_i(\omega)$  by A.

For  $\Phi \subseteq \text{At}$  with  $S_\Phi \preceq \alpha_i(\omega)$ ,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_\Phi$  by B. It follows that  $\Pi_i^*(\omega)_\Phi = \Pi_i^*(\omega_\Phi)$ . Hence  $\Pi_i^*(\omega)_\Upsilon = \Pi_i^*(\omega_\Upsilon)$  for  $\Upsilon \subseteq \Phi$ .

For  $\Phi \subseteq \text{At}$  with  $S_\omega \succeq S_\Phi \succeq \alpha_i(\omega)$ ,  $\Pi_i^*(\omega_\Phi) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$  by C. Hence  $\Pi_i^*(\omega_\Phi) = \Pi_i^*(\omega)$ . It follows that  $\Pi_i^*(\omega_\Phi)_\Upsilon = \Pi_i^*(\omega)_\Upsilon = \Pi_i^*(\omega_\Upsilon)$  for  $S_\Upsilon \preceq \alpha_i(\omega)$ .  $\square$

We conclude that the derived explicit possibility correspondence  $\Pi_i^*$  is a possibility correspondence as in Heifetz, Meier, and Schipper (2006, 2008), i.e., satisfies Assumption 1. To show that the connection between the derived explicit possibility correspondence and the implicit possibility correspondence is as in the complemented HMS model of the prior section, we note the following lemma.

**Lemma 13** *For any individual  $i \in I$ ,  $\Lambda_i^*$  and  $\Pi_i^*$  jointly satisfy explicit and implicit measurability.*

PROOF. For explicit measurability, we need to show that for all  $\omega \in \Omega$ ,  $\omega' \in \Lambda_i^*(\omega)$  implies  $\Pi_i^*(\omega') = \Pi_i(\omega)$ . By Stationarity of  $\Lambda_i^*$ ,  $\omega' \in \Lambda_i^*(\omega)$  implies  $\Lambda_i^*(\omega') = \Lambda_i^*(\omega)$ . By Lemma 7 A., we have  $\Pi_i^*(\omega') = \Lambda_i^*(\omega')_{\alpha_i(\omega')}$  and  $\Pi_i^*(\omega) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$ . By Awareness Measurability,  $\omega' \in \Lambda_i^*(\omega)$  implies  $\alpha_i(\omega') = \alpha_i(\omega)$ . Thus,  $\Pi_i^*(\omega') = \Pi_i(\omega)$ .

For implicit measurability, we need to show that for all  $\omega \in \Omega$ ,  $\omega' \in \Pi_i^*(\omega)$ ,  $\Lambda_i^*(\omega') = \Lambda_i^*(\omega)_{S_{\Pi_i^*(\omega)}}$ . By Lemma 10 (Stationarity),  $\omega' \in \Pi_i^*(\omega)$ ,  $\Pi_i^*(\omega') = \Pi_i^*(\omega)$ . This implies that  $S_{\omega'} = S_{\Pi_i^*(\omega)}$ . Since by Lemma 7 A., we have  $\Pi_i^*(\omega') = \Lambda_i^*(\omega')_{\alpha_i(\omega')}$  and  $\Pi_i^*(\omega) = \Lambda_i^*(\omega)_{\alpha_i(\omega)}$ , it follows now  $\alpha_i(\omega') = \alpha_i(\omega) = S_{\Pi_i^*(\omega)}$ . From Strong Confinement of  $\Lambda_i^*$  (implied by Reflexivity and Stationarity of  $\Lambda_i^*$ ) follows that  $\omega' \in \Lambda_i^*(\omega')$  and hence  $\Lambda_i^*(\omega') \subseteq S_{\Pi_i^*(\omega)}$ . It follows now that  $\Lambda_i^*(\omega') = \Pi_i^*(\omega') = \Pi_i^*(\omega) = \Lambda_i^*(\omega)_{S_{\Pi_i^*(\omega)}}$ .  $\square$

The above lemmata imply the following:

**Corollary 1** *For any implicit knowledge-based HMS model  $\mathbf{M}^* = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  with derived explicit possibility correspondences  $(\Pi_i^*)_{i \in I}$  we have that  $\bar{\mathbf{M}} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, v \rangle$  is a complemented HMS model and  $\mathbf{M} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Pi_i^*)_{i \in I}, v \rangle$  is a HMS model.*

The awareness function can be directly used to define an awareness operator on events.

**Definition 9 (Awareness Operator II)** *For each individual  $i \in I$ , define an awareness operator on events by for all  $E \in \Sigma$ ,*

$$A_i^*(E) := \{\omega \in \Omega : \alpha_i(\omega) \succeq S(E)\}$$

*if there is a state  $\omega \in \Omega$  such that  $\alpha_i(\omega) \succeq S(E)$  and by  $A_i^*(E) = \emptyset^{S(E)}$  otherwise.*

Similarly, for each individual  $i \in I$ , we can use the possibility correspondence  $\Lambda_i^*$  to define an implicit knowledge operator  $L_i^*$  as in Definition 6. Finally, let  $K_i$  be the explicit knowledge operator and  $A_i$  be the awareness operator defined from the derived explicit possibility correspondence  $\Pi_i^*$  as in Definitions 2 and 3, respectively.

The following proposition shows that awareness defined with the awareness function is equivalent to awareness defined with the derived explicit possibility correspondence. It also shows that explicit knowledge defined from the derived explicit possibility correspondence is equivalent to implicit knowledge and awareness defined from the primitive implicit possibility correspondence and awareness function.

**Proposition 5** *For every individual  $i \in I$  and any event  $E \in \Sigma$ ,*

1.  $A_i^*(E) = A_i(E)$

$$2. K_i(E) = L_i^*(E) \cap A_i^*(E).$$

PROOF. 1. For any  $E \in \Sigma$ ,  $\omega \in A_i^*(E)$  if and only if  $\alpha_i(\omega) \succeq S(E)$ . By definition of  $\Pi_i^*$ , this holds if and only if  $\Pi_i^*(\omega) \subseteq S \succeq S(E)$ . This is equivalent to  $\omega \in A_i(E)$ .

2. For any non-empty  $E \in \Sigma$ ,  $\omega \in K_i(E)$  if and only if  $\Pi_i^*(\omega) \subseteq E$ . By definition of  $\Pi_i^*$ , it holds if and only if  $\Lambda_i^*(\omega) \subseteq E$  and  $\alpha_i(\omega) \succeq S(E)$ . This is equivalent to  $\omega \in L_i^*(E)$  and  $\omega \in A_i^*(E)$ . If  $E = \emptyset^{S_\Phi}$  for some  $\Phi \subseteq \text{At}$ , then  $K_i(E) = \emptyset^{S_\Phi}$  and  $L_i^*(E) = \emptyset^{S_\Phi}$ .  $A_i^*(E) \subseteq S_\Phi^\uparrow$ . Thus, by conjunction of events  $L_i^*(E) \cap A_i^*(E) = \emptyset^{S_\Phi}$ .  $\square$

Together, Sections 3 and 4 show an interdefinability of explicit and implicit knowledge in HMS models. Implicit knowledge can be defined in terms of explicit knowledge and explicit knowledge can be defined in terms of implicit knowledge. We can use either the explicit possibility correspondences as primitive or the implicit possibility correspondence together with the awareness function. Implicit knowledge-based HMS models are arguably closer to FH models than HMS models. We will use them to build a bridge between HMS and FH models.

## 5 Category of FH Models

In this section, we introduce FH models and bounded morphisms, a notion of structure preserving maps between FH models, and use these to form a category with FH models as objects and bounded morphisms as morphisms. As mentioned in the introduction, the goal is to provide subjective versions of FH models.

The semantics of FH models is not syntax-free since each agent's awareness function assigns to each state a set of formulas. Thus, we first introduce the formal language featuring implicit knowledge, awareness, and explicit knowledge. With  $i \in I$  and  $p \in \text{At}$ , define the language  $\mathcal{L}_{\text{At}}$  by

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \ell_i\varphi \mid a_i\varphi \mid k_i\varphi$$

With some abuse of notation, let  $\text{At}(\varphi) := \{p \in \text{At} : p \text{ is a subformula of } \varphi\}$ , for any  $\varphi \in \mathcal{L}_{\text{At}}$ , be the set of atomic formulas that are contained in  $\varphi$ , and let  $\mathcal{L}_\Phi := \{\varphi \in \mathcal{L}_{\text{At}} : \text{At}(\varphi) \subseteq \Phi\}$  be the sublanguage of  $\mathcal{L}_{\text{At}}$  built on propositions  $p$  in  $\Phi \subseteq \text{At}$ .

The formula  $\ell_i\varphi$  reads “agent  $i$  implicitly knows  $\varphi$ ”,  $a_i\varphi$  reads “ $i$  is aware of  $\varphi$ ”, and  $k_i\varphi$  reads “ $i$  explicitly knows  $\varphi$ ”. Fagin and Halpern (1988) define explicit knowledge as the conjunction of implicit knowledge and awareness, namely  $k_i\varphi = a_i\varphi \wedge \ell_i\varphi$ , for  $\varphi \in \mathcal{L}_{\text{At}}$ .

**Definition 10 (FH Model)** *For any  $\Phi \subseteq \text{At}$ , an FH model  $\mathbf{K}_\Phi = \langle I, W_\Phi, (R_{\Phi,i})_{i \in I}, (\mathcal{A}_{\Phi,i})_{i \in I}, V_\Phi \rangle$  for  $\Phi$  consists of*

- a non-empty set of individuals  $I$ ,

- a non-empty set of states  $W_\Phi$ ,
- an accessibility relation  $R_{\Phi,i} \subseteq W_\Phi \times W_\Phi$ , for all  $i \in I$ ,
- an awareness function  $\mathcal{A}_{\Phi,i} : W_\Phi \longrightarrow 2^{\mathcal{L}_\Phi}$ , for all  $i \in I$ , assigning to each state  $w \in W_\Phi$  a set of formulas  $\mathcal{A}_{\Phi,i}(w) \subseteq \mathcal{L}_\Phi$ . The set  $\mathcal{A}_{\Phi,i}(w)$  is called the awareness set of  $i$  at  $w$ .
- a valuation function  $V_\Phi : \Phi \longrightarrow 2^{W_\Phi}$ .

**Assumption 4 (Properties Imposed on FH Models)** *We require that the FH model  $K_\Phi$  is propositionally determined, i.e., for every  $i \in I$ , the awareness functions satisfy<sup>33</sup>*

Awareness is Generated by Primitive Propositions: *For all  $\varphi \in \mathcal{L}_\Phi$ ,  $\varphi \in \mathcal{A}_{\Phi,i}(w)$  if and only if for all  $p \in \text{At}(\varphi)$ ,  $p \in \mathcal{A}_{\Phi,i}(w)$ .*

Agents Know What They Are Aware of:  *$(w, t) \in R_{\Phi,i}$  implies  $\mathcal{A}_{\Phi,i}(w) = \mathcal{A}_{\Phi,i}(t)$ .*

*We also require that the FH model  $K_\Phi$  is partitional, that is,  $R_{\Phi,i}$  is an equivalence relation, i.e., satisfies reflexivity, transitivity, and Euclideaness.*

Throughout the paper, we focus on partitional and propositionally determined FH models because these models capture the notion of awareness and knowledge used in most applications so far and it is also the notion of awareness used in HMS models. We are interested in how this rich structure maps between FH models as well as between FH and HMS models.

**Definition 11 (Bounded Morphism)** *For any  $\Psi \subseteq \Phi \subseteq \text{At}$  and FH models  $K_\Phi = \langle I, W_\Phi, (R_{\Phi,i})_{i \in I}, (\mathcal{A}_{\Phi,i})_{i \in I}, V_\Phi \rangle$  and  $K_\Psi = \langle I, W_\Psi, (R_{\Psi,i})_{i \in I}, (\mathcal{A}_{\Psi,i})_{i \in I}, V_\Psi \rangle$ , the mapping  $f_\Psi^\Phi : K_\Phi \longrightarrow K_\Psi$  is a surjective bounded morphism if for every  $i \in I$  and  $w \in W_\Phi$*

- Surjectivity:  $f_\Psi^\Phi : W_\Phi \longrightarrow W_\Psi$  is a surjection,
- Atomic harmony: *for every  $p \in \Psi$ ,  $w \in V_\Phi(p)$  if and only if  $f_\Psi^\Phi(w) \in V_\Psi(p)$ ,*
- Awareness consistency:  $\mathcal{A}_{\Phi,i}(w) \cap \mathcal{L}_\Psi = \mathcal{A}_{\Psi,i}(f_\Psi^\Phi(w))$ ,
- Homomorphism:  $f_\Psi^\Phi$  is a homomorphism with respect to  $R_{\Phi,i}$ , i.e., if  $(w, t) \in R_{\Phi,i}$ , then  $(f_\Psi^\Phi(w), f_\Psi^\Phi(t)) \in R_{\Psi,i}$ ,
- Back: *if  $(f_\Psi^\Phi(w), t') \in R_{\Psi,i}$ , then there is a state  $t \in W_\Phi$  such that  $f_\Psi^\Phi(t) = t'$  and  $(w, t) \in R_{\Phi,i}$ .*

---

<sup>33</sup>The terminology is from Halpern (2001).

This is the standard notion of surjective bounded morphism (also called surjective  $p$ -morphism) (see for instance, Blackburn, de Rijke, and Venema, 2001, pp. 59–62) except for the additional property of Awareness Consistency. In our context, the bounded morphism literally bounds the language over which FH models are defined. We can now consider collections of FH models and commuting bounded morphisms between them as follows:

**Definition 12 (Category of FH Models)** *Given the FH model  $K_{At}$ , the category of FH models  $\mathcal{C}(K_{At}) = \langle (K_\Phi)_{\Phi \subseteq At}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq At} \rangle$  consists of*

- *a collection of FH models  $K_\Phi$ , one for each  $\Phi \subseteq At$ ,*
- *for any  $\Phi, \Psi \subseteq At$  with  $\Psi \subseteq \Phi$ , there is one surjective bounded morphism  $f_\Psi^\Phi$ , such that*
  - *for any  $\Phi \subseteq At$ ,  $f_\Phi^\Phi$  is the identity,*
  - *for any  $\Upsilon, \Phi, \Psi \subseteq At$  with  $\Upsilon \subseteq \Psi \subseteq \Phi$ ,  $f_\Upsilon^\Phi = f_\Upsilon^\Psi \circ f_\Psi^\Phi$ .*

Our terminology is not arbitrary. The category of FH models is indeed a category in the sense of category theory. It has an initial object, the most expressive FH model  $K_{At}$ , as well as a terminal object, the FH model  $K_\emptyset$ .

Since the category of FH models is defined with bounded morphisms, it suggests that all FH models in the category are in some sense epistemically equivalent. Indeed, we interpret each category of FH models literally as the category of FH models that vary with the language but are otherwise modally equivalent. That is, for any  $\Psi \subseteq \Phi \subseteq At$ , modal satisfaction for  $K_\Psi$  is as for  $K_\Phi$  with respect to formulas in  $\mathcal{L}_\Psi$  (see Lemma 14 below). We interpret this as follows: When a modeler represents a context with a FH model  $K_{At}$ , an agent  $i$  at state  $w \in W_{At}$  can be thought of representing it with the FH model  $K_{At(\mathcal{A}_{At,i}(w))}$ .<sup>34</sup> And this agent  $i$  considers it possible at  $w$  that at  $t$  with  $(f_{At(\mathcal{A}_{At,i}(w))}^{At}(w), t) \in R_{At(\mathcal{A}_{At,i}(w)),i}$  agent  $j$  represents the situation with the FH model  $K_{At(\mathcal{A}_{At,j}(t))}$ , etc. These models can all be seen as equivalent except for the language of which they are defined. With this construction, we do not just endow agents with a formal language to reason about their context but we also allow them to analyze their context with semantic devices like logicians do. This is relevant because in many multi-agent contexts of game theory, the analysis proceeds using semantic devices like state spaces etc. rather than at the level of syntax. For instance, in a principal-agent problem, the principal may want to use a FH model augmented by actions and utility functions to analyze optimal contract design realizing that the (unaware) agent may also use a less expressive but otherwise equivalent FH model to analyze how to optimally interact with the principal.

To make the equivalence between models in the category of FH models precise, we need to introduce the semantics of FH models.

<sup>34</sup>As usual,  $At(\mathcal{A}_{At,i}(w))$  denotes the image under the function  $At(\cdot)$  of  $\mathcal{A}_{At,i}(w)$ .

**Definition 13 (FH Semantics)** For any  $\Phi \subseteq \text{At}$ , FH model  $\mathbf{K}_\Phi = \langle I, W_\Phi, (R_{\Phi,i})_{i \in I}, (\mathcal{A}_{\Phi,i})_{i \in I}, V_\Phi \rangle$ , and  $\omega \in W_\Phi$ , satisfaction of formulas in  $\mathcal{L}_\Phi$  is given by the following clauses:

$$\begin{array}{llll} \mathbf{K}_\Phi, w \Vdash \top & \text{for all } w \in W_\Phi; & \mathbf{K}_\Phi, w \Vdash \varphi \wedge \psi & \text{iff } \mathbf{K}_\Phi, w \Vdash \varphi \text{ and } \mathbf{K}_\Phi, w \Vdash \psi; \\ \mathbf{K}_\Phi, w \Vdash p & \text{iff } w \in V_\Phi(p); & \mathbf{K}_\Phi, w \Vdash a_i \varphi & \text{iff } \varphi \in \mathcal{A}_{\Phi,i}(w); \\ \mathbf{K}_\Phi, w \Vdash \neg \varphi & \text{iff } \mathbf{K}_\Phi, w \not\Vdash \varphi; & \mathbf{K}_\Phi, w \Vdash \ell_i \varphi & \text{iff } \mathbf{K}_\Phi, t \Vdash \varphi \text{ for all } (w, t) \in R_{\Phi,i}. \end{array}$$

From this semantics and the syntactic definition  $k_i \varphi := \ell_i \varphi \wedge a_i \varphi$ , it follows that  $\mathbf{K}_\Phi, w \Vdash k_i \varphi$  if and only if for all  $t$  such that  $(w, t) \in R_{\Phi,i}$ ,  $\mathbf{K}_\Phi, t \Vdash \varphi$  and  $\varphi \in \mathcal{A}_{\Phi,i}(w)$ .

The category of FH models forms a complete lattice induced by set inclusion on sets of atomic formulas with the initial object being the meet of the lattice and the terminal object being the join of the lattice. We now show that it gives rise to a complete lattice when ordered using the (directed) bounded morphism or, epistemically more relevant, when ordered by modal equivalence relative to sublanguages. To prove the proposition showing that a category forms a complete lattice, we need the following lemma, which shows that for any two FH models in the category, they are modally equivalent with respect to the language for which the less expressive FH models is defined.

**Lemma 14** Given a category of FH models,  $\langle (\mathbf{K}_\Phi)_{\Phi \subseteq \text{At}}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}} \rangle$ , for any  $\Psi, \Phi \subseteq \text{At}$  with  $\Psi \subseteq \Phi$ , all  $w \in W_\Phi$ , and all  $\varphi \in \mathcal{L}_\Psi$ ,

$$\mathbf{K}_\Phi, w \Vdash \varphi \text{ if and only if } \mathbf{K}_\Psi, f_\Psi^\Phi(w) \Vdash \varphi.$$

**PROOF.** The proof is by induction on the complexity of formulas in  $\mathcal{L}_\Psi$ . The base case follows directly from Atomic harmony. The Boolean cases are straightforward once we note that these formulas are restricted to  $\mathcal{L}_\Psi$ .

Now consider the case  $\ell_i \varphi \in \mathcal{L}_\Psi$ :

“ $\Rightarrow$ ”: Assume  $\mathbf{K}_\Phi, w \Vdash \ell_i \varphi$ . This holds if and only if  $\mathbf{K}_\Phi, w' \Vdash \varphi$  for all  $(w, w') \in R_{\Phi,i}$ . By the induction hypothesis,  $\mathbf{K}_\Psi, f_\Psi^\Phi(w') \Vdash \varphi$ . Since  $f_\Psi^\Phi$  is a homomorphism,  $(f_\Psi^\Phi(w), f_\Psi^\Phi(w')) \in R_{\Psi,i}$ . Suppose by contradiction that there exists  $w'' \in W_\Psi$  such that  $\mathbf{K}_\Psi, w'' \not\Vdash \varphi$  but  $(f_\Psi^\Phi(w), w'') \in R_{\Psi,i}$ . By the Back condition, there exists  $w''' \in W_\Phi$  such that  $f_\Psi^\Phi(w''') = w''$  and  $(w, w''') \in R_{\Phi,i}$ . By Atomic harmony,  $\mathbf{K}_\Phi, w''' \not\Vdash \varphi$ . This contradicts now  $\mathbf{K}_\Phi, w' \Vdash \varphi$  for all  $(w, w') \in R_{\Phi,i}$ .

“ $\Leftarrow$ ”: Assume  $\mathbf{K}_\Psi, f_\Psi^\Phi(w) \Vdash \ell_i \varphi$ . This holds if and only if  $\mathbf{K}_\Psi, w' \Vdash \varphi$  for all  $(f_\Psi^\Phi(w), w') \in R_{\Psi,i}$ . Suppose by contradiction that there exists  $w'' \in (f_\Psi^\Phi)^{-1}(w)$  such that  $\mathbf{K}_\Phi, w'' \not\Vdash \varphi$  for some  $w'' \in W_\Phi$  with  $(w'', w') \in R_{\Phi,i}$ . By the induction hypothesis,  $\mathbf{K}_\Psi, f_\Psi^\Phi(w'') \not\Vdash \varphi$ . Since  $f_\Psi^\Phi$  is a homomorphism,  $(f_\Psi^\Phi(w''), f_\Psi^\Phi(w')) \in R_{\Psi,i}$ . But  $f_\Psi^\Phi(w'') = f_\Psi^\Phi(w)$ . Thus, we have a contradiction to  $\mathbf{K}_\Psi, w' \Vdash \varphi$  for all  $(f_\Psi^\Phi(w), w') \in R_{\Psi,i}$ .

Now consider the case  $a_i \varphi \in \mathcal{L}_\Psi$ :

We have that  $\mathbf{K}_\Phi, w \Vdash a_i \varphi$  if and only if  $\varphi \in \mathcal{A}_{\Phi,i}(w)$  by the definition of semantics. By Awareness consistency,  $\mathcal{A}_{\Psi,i}(f_\Psi^\Phi(w)) = \mathcal{A}_{\Phi,i}(w) \cap \mathcal{L}_\Psi$ . Moreover,  $a_i \varphi \in \mathcal{L}_\Psi$  implies  $\varphi \in \mathcal{L}_\Psi$ . Thus,  $\varphi \in \mathcal{A}_{\Psi,i}(f_\Psi^\Phi(w))$ . This holds if and only if  $\mathbf{K}_\Psi, f_\Psi^\Phi(w) \Vdash a_i \varphi$  (by the

definition of the semantics).

□

We are now ready to claim that the category of FH models is a complete lattice when ordered by modal equivalence modulo sublanguages.

**Proposition 6** *Given a category of FH models,  $\langle (K_\Phi)_{\Phi \subseteq \text{At}}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}} \rangle$ , modal equivalence relative to sublanguages forms a complete lattice of FH models in the category as follows: For any nonempty set of subsets of atomic formulas  $\mathcal{F} \subseteq 2^{\text{At}}$ ,*

- (i)  $K_{\bigcup_{\Phi \in \mathcal{F}} \Phi}$  is modally equivalent to  $K_\Psi$  w.r.t.  $\mathcal{L}_\Psi$  for every  $\Psi \in \mathcal{F}$ , i.e., for any  $w \in W_{\bigcup_{\Phi \in \mathcal{F}} \Phi}$ ,  $\varphi \in \mathcal{L}_\Psi$ ,  $K_{\bigcup_{\Phi \in \mathcal{F}} \Phi}, w \Vdash \varphi$  iff  $K_\Psi, f_{\bigcup_{\Phi \in \mathcal{F}} \Phi}^\Psi(w) \Vdash \varphi$ , and
- (ii)  $K_{\bigcap_{\Phi \in \mathcal{F}} \Phi}$  is modally equivalent to  $K_\Psi$  w.r.t.  $\mathcal{L}_{\bigcap_{\Phi \in \mathcal{F}} \Phi}$  for every  $\Psi \in \mathcal{F}$ , i.e., for any  $w \in W_\Psi$ ,  $\varphi \in \mathcal{L}_{\bigcap_{\Phi \in \mathcal{F}} \Phi}$ ,  $K_\Psi, w \Vdash \varphi$  iff  $K_{\bigcap_{\Phi \in \mathcal{F}} \Phi}, f_{\bigcap_{\Phi \in \mathcal{F}} \Phi}^\Psi(w) \Vdash \varphi$ .

PROOF. By Lemma 14, modal equivalence w.r.t. sublanguages is an order on objects in the category of FH models. To show that modal equivalence w.r.t. sublanguages forms a complete lattice of FH models, notice that reflexivity follows from the requirement of the FH category that the bounded morphism from the FH model to itself is the identity. Antisymmetry follows from the requirement of the FH category that if there is a bounded morphism from one FH model to another and back, then it must be identity. Transitivity follows from composition of bounded morphisms. To show that for every subset of FH models there is a join and a meet, we need to prove (i) and (ii) respectively. To prove (i), let  $\Phi = \bigcup_{\Upsilon \in \mathcal{F}} \Upsilon$  and  $\Psi \in \mathcal{F}$  and apply Lemma 14. To prove (ii), let  $\Psi = \bigcap_{\Upsilon \in \mathcal{F}} \Upsilon$  and  $\Phi \in \mathcal{F}$  and apply Lemma 14. This completes the proof of the proposition. □

Note that for a collection of FH models  $\{K_\Psi\}_{\Psi \in \mathcal{F}}$ , the “join” and “meet” FH models are  $K_{\bigcup_{\Psi \in \mathcal{F}} \Psi}$  and  $K_{\bigcap_{\Psi \in \mathcal{F}} \Psi}$ , respectively. So for any collection of FH models, Proposition 6 shows modal equivalence between any FH model in the collection and its join and meet models, respectively.

Our notion of bounded morphism is inspired by bisimulation of FH models introduced by van Ditmarsch et al. (2018). Clearly, the surjective bounded morphism is a bisimulation (Blackburn et al., 2001). Here we discuss some differences and similarities between the notions. While bisimulation more generally is a relation between models without a particular direction, the bounded morphism has a natural direction from the more expressive FH model to the less expressive FH model. Further, it is a function on  $W_\Phi$ . That is, it maps *every* state in  $W_\Phi$  to a state in  $W_\Psi$  with  $\Psi \subseteq \Phi$ . Moreover, surjectivity is a property that is straightforward to define for functions. Finally, bounded morphisms easily compose and almost naturally lead to the notion of category of FH models although we do not really make much use here of the machinery of category theory.<sup>35</sup> For all these reasons, we use the notion of bounded morphism. Van Ditmarsch

<sup>35</sup>There is a growing literature making use of categories of Kripke frames with bounded morphisms between them, e.g., Moss (1999).

et al. (2018) introduced two notions of bisimulation for FH models, standard bisimulation and awareness bisimulation. Like our bounded morphism, both of their notions of bisimulation also depend on a subset of atomic formulas for FH models. The clauses Atomic harmony, Awareness consistency, Homomorphism, and Back have counterparts in their notions of bisimulations. Our notion of bounded morphism is closer to what they call standard bisimulation because our Homomorphism and Back clauses do not involve the awareness function. Although van Ditmarsch et al. (2018, p. 63) mention the projective lattice structure of Heifetz, Meier, and Schipper (2006, 2008) as a motivation for their notion of awareness bisimulation, we believe it is particularly useful for their notion of speculative knowledge. Their notions of bisimulations do not explicitly require surjectivity. However, when they consider maximal bisimulations, they must be automatically surjective since they yield quotient models. Compositions of maximal bisimulation commute like we require our bounded morphism to do in the category of FH models but bisimulations that are not maximal do not necessarily commute. Moreover, maximal bisimulations yield necessarily contractions that eliminate redundancies.<sup>36</sup> We are unsure of whether it is necessarily a natural property when we interpret FH models as subjective views of agents. An agent may not realize or may not be bothered by redundancies and use an FH model with redundancies to analyze her situation. That is, differences in awareness and redundancies are orthogonal to each other and reduction in awareness does not necessitate elimination of redundancies.

## 6 Transformations

### 6.1 From FH Models to Complemented HMS Models

We can use the tools of the prior sections to define a transformation of a FH model into a complemented HMS model. The transformation works as follows: For any FH model  $K_{At}$  for  $At$ , consider the category of FH models  $\langle (K_\Phi)_{\Phi \subseteq At}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq At} \rangle$ . This category is transformed into an implicit knowledge-based HMS model. We then derive the explicit possibility correspondences and add them to the implicit knowledge-based HMS model, obtaining a complemented implicit knowledge-based HMS model as in Section 4. In the next step, we erase the awareness functions and get a complemented HMS model. The core step is to transform a category of FH models into an implicit knowledge-based HMS model. This is defined next.

**Definition 14 (*T-Transform*)** *For any category of FH models  $\mathcal{C}(K_{At}) = \langle (K_\Phi)_{\Phi \subseteq At}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq At} \rangle$ , the *T-transform* of  $\mathcal{C}(K_{At})$  is the implicit knowledge-based HMS model  $T(\mathcal{C}(K_{At})) = \langle I, \{S_\Phi\}_{\Phi \subseteq At}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq At}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  defined as follows:*

- $S_\Phi = W_\Phi$  for all  $\Phi \subseteq At$ , where  $W_\Phi$  is the state space of the FH model  $K_\Phi$  of the category  $\mathcal{C}(K_{At})$ . Denote  $\Omega = \bigcup_{\Phi \subseteq A} S_\Phi$ .

---

<sup>36</sup>Also, contractions do not yield disjoint state spaces. To obtain disjoint state spaces, we would have to consider additionally models isomorphic to contractions but with disjoint state spaces.



- $r_\Psi^\Phi = f_\Psi^\Phi$  for any  $\Phi, \Psi \subseteq \text{At}$  with  $\Psi \subseteq \Phi$ , where  $f_\Psi^\Phi$  is the surjective bounded morphism of the category  $\mathcal{C}(\mathbf{K}_{\text{At}})$ .
- $\Lambda_i^* : \Omega \longrightarrow 2^\Omega$  such that  $\Phi \subseteq \text{At}$  and  $w \in S_\Phi$ ,  $w' \in \Lambda_i^*(w)$  if and only if  $(w, w') \in R_{\Phi,i}$ , for any  $i \in I$ .
- $\alpha_i : \Omega \longrightarrow \{S_\Phi\}_{\Phi \subseteq \text{At}}$  such that for all  $\Psi \subseteq \text{At}$  and  $w \in S_\Psi$ ,  $\alpha_i(w) = S_\Upsilon$  if and only if  $\text{At}(\mathcal{A}_{\Psi,i}(w)) = \Upsilon$ , for any  $i \in I$ .
- $v(p) = \bigcup_{\Phi \subseteq \text{At}} V_\Phi(p)$ , for any  $p \in \text{At}$ .

The  $T$ -transform indeed transforms any category of  $FH$  models into an implicit knowledge-based HMS model.

**Proposition 7** *For any category of  $FH$  models  $\mathcal{C}(\mathbf{K}_{\text{At}})$ , the  $T$ -transform  $T(\mathcal{C}(\mathbf{K}_{\text{At}}))$  is an implicit knowledge-based HMS model.*

PROOF. Fix a category of  $FH$  models,  $\mathcal{C}(\mathbf{K}_{\text{At}}) = \langle (\mathbf{K}_\Phi)_{\Phi \subseteq \text{At}}, (f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}} \rangle$  with  $\mathbf{K}_\Phi = \langle I, W_\Phi, (R_{\Phi,i})_{i \in I}, (\mathcal{A}_{\Phi,i})_{i \in I}, V_\Phi \rangle$  for  $\Phi \subseteq \text{At}$ . We show that we can obtain from this category each component of an implicit knowledge-based HMS model.

*Collection of state spaces:*  $\{S_\Phi\}_{\Phi \subseteq \text{At}}$  forms a complete lattice by set inclusion on  $\text{At}$ .

*Projections:* Since  $r_\Psi^\Phi = f_\Psi^\Phi$  for any  $\Phi, \Psi \subseteq \text{At}$  with  $\Psi \subseteq \Phi$ , the fact that projections commute follows immediately from  $(f_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}$  being morphisms of the category  $\mathcal{C}(\mathbf{K}_{\text{At}})$ . It also implies that  $r_\Phi^\Phi$  is the identity on  $S_\Phi$ . The fact that  $r_\Psi^\Phi$  is a surjection follows directly from  $f_\Psi^\Phi$  being a surjective bounded morphism.

*Implicit possibility correspondences:* For every  $\Phi \subseteq \text{At}$ , the partitionial properties of  $\Lambda_i^*$  restricted to  $S_\Phi$  follow directly from the partial properties of  $R_{\Phi,i}$ . We need to show Projections Preserve Implicit Knowledge, that is, we need to show that for any  $\Phi \subseteq \text{At}$ , if  $\omega \in S_\Phi$ , then  $\Lambda_i^*(\omega)_\Psi = \Lambda_i^*(\omega_\Psi)$ .

“ $\subseteq$ ”: If  $\omega' \in \Lambda_i^*(\omega)_\Psi$ , then  $(f_\Psi^\Phi)^{-1}(\omega') \subseteq \Lambda_i^*(\omega)$  since  $f_\Psi^\Phi$  is a surjection. For all  $\omega'' \in (f_\Psi^\Phi)^{-1}(\omega')$ ,  $(\omega, \omega'') \in R_{\Phi,i}$  by  $T$ -transform. Since  $f_\Psi^\Phi$  is a homomorphism,  $(f_\Psi^\Phi(\omega), f_\Psi^\Phi(\omega'')) \in R_{\Psi,i}$ . By the  $T$ -transform,  $f_\Psi^\Phi(\omega'') \in \Lambda_i^*(\omega_\Psi)$ . Thus,  $\omega' \in \Lambda_i^*(\omega_\Psi)$ .

“ $\supseteq$ ”: If  $\omega' \in \Lambda_i^*(\omega_\Psi)$ , then by the  $T$ -transform  $(\omega_\Psi, \omega') \in R_{\Psi,i}$ . By Back, there exists  $t \in (f_\Psi^\Phi)^{-1}(\omega')$  with  $(\omega, t) \in R_{\Phi,i}$ . By the  $T$ -transform,  $t \in \Lambda_i^*(\omega)$ . Hence  $\omega' \in \Lambda_i^*(\omega)_\Psi$ .

*Awareness function:* We verify the properties one-by-one.

0.: For all  $w \in W_\Phi$ ,  $\mathcal{A}_{\Phi,i}(w) \subseteq \mathcal{L}_\Phi$ . Thus,  $\text{At}(\mathcal{A}_{\Phi,i}(w)) \subseteq \Phi$ . Set  $\Psi = \text{At}(\mathcal{A}_{\Phi,i}(w))$ . By  $T$ -transform,  $\alpha_i(w) = S_\Psi$ .

I.: If  $\omega' \in \Lambda_i^*(\omega)$  then by the  $T$ -transform  $(\omega, \omega') \in R_{\Phi,i}$  for  $\Phi \subseteq \text{At}$  with  $\omega' \in S_\Phi$ . Since in a propositionally determined FH model, agents know what they are aware of,  $\mathcal{A}_{\Phi,i}(\omega) = \mathcal{A}_{\Phi,i}(\omega')$ . By the  $T$ -transform,  $\alpha_i(\omega) = \alpha_i(\omega')$ .

II.: If  $\omega \in S_\Phi$  and  $S_\Psi \preceq \alpha_i(\omega)$ , then by the  $T$ -transform  $\text{At}(\mathcal{A}_{\Phi,i}(\omega)) \supseteq \Psi$ . By Awareness consistency,  $\mathcal{A}_{\Psi,i}(f_\Psi^\Phi(\omega)) \cap \mathcal{L}_\Psi$ . Thus,  $\text{At}(\mathcal{A}_{\Psi,i}(\omega_\Psi)) = \Psi$ . By the  $T$ -transform,  $\alpha_i(\omega_\Psi) = S_\Psi$ .

III.: If  $\omega \in S_\Phi$  and  $\alpha_i(\omega) \preceq S_\Psi \preceq S_\Phi$ , then by the  $T$ -transform  $\text{At}(\mathcal{A}_{\Phi,i}(\omega)) \subseteq \Psi \subseteq \Phi$ . Denote  $\Upsilon = \text{At}(\mathcal{A}_{\Phi,i}(\omega))$ . By Awareness consistency,  $\mathcal{A}_{\Psi,i}(f_\Psi^\Phi(\omega)) = \mathcal{A}_{\Phi,i}(\omega) \cap \mathcal{L}_\Psi = \mathcal{L}_\Upsilon$ . Thus,  $\text{At}(\mathcal{A}_{\Psi,i}(f_\Psi^\Phi(\omega))) = \Upsilon$ . By the  $T$ -transform,  $\alpha_i(\omega_\Psi) = \alpha_i(\omega)$ .

IV.: If  $\omega \in S_\Phi$  and  $\Psi \subseteq \Phi$ , then by Awareness consistency  $\mathcal{A}_{\Phi,i}(\omega) \cap \mathcal{L}_\Psi = \mathcal{A}_{\Psi,i}(f_\Psi^\Phi(\omega))$ . Thus,  $\text{At}(\mathcal{A}_{\Phi,i}(\omega)) \supseteq \text{At}(\mathcal{A}_{\Psi,i}(f_\Psi^\Phi(\omega)))$ . Applying the  $T$ -transform, yields  $\alpha_i(\omega) \succeq \alpha_i(\omega_\Psi)$ .

*Valuation:* We have to show that  $v(p)$  is an event for every  $p \in \text{At}$ . Take  $S_{\{p\}}$  to be the base-space and  $V_{\{p\}}(p)$  to be the base. Then it follows immediately from the definition of  $T$ -transform that  $v(p)$  is an event.

This completes the proof of the proposition.  $\square$

We have all ingredients to define the transformation of FH models into complemented HMS models.

**Definition 15 (HMS-Transform)** *For any FH model  $\mathbf{K}_{\text{At}}$ , the HMS-transform  $HMS(\mathbf{K}_{\text{At}}) = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, v \rangle$  is defined by the following steps:*

1. *Form the category of FH models  $\mathcal{C}(\mathbf{K}_{\text{At}})$  (Definition 12).*
2. *Apply the  $T$ -transform to  $\mathcal{C}(\mathbf{K}_{\text{At}})$  to obtain the implicit knowledge-based HMS model  $T(\mathcal{C}(\mathbf{K}_{\text{At}})) = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  (Definition 14).*
3. *Form the complemented implicit knowledge-based HMS model  $\overline{T(\mathcal{C}(\mathbf{K}_{\text{At}}))} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  by deriving the explicit possibility correspondences  $(\Pi_i^*)_{i \in I}$  (Definition 8).*
4. *Erase the awareness functions  $(\alpha_i)_{i \in I}$  from the complemented implicit knowledge-based HMS model  $\overline{T(\mathcal{C}(\mathbf{K}_{\text{At}}))}$  to obtain the complemented HMS model  $\langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, v \rangle$ .*

**Corollary 2** *For any FH model  $\mathbf{K}_{\text{At}}$ , its HMS-transform  $HMS(\mathbf{K}_{\text{At}})$  is a complemented HMS model.*

**PROOF.** The proof is a corollary from the preceding results. For any  $\mathbf{K}_{\text{At}}$ , consider the category  $\mathcal{C}(\mathbf{K}_{\text{At}})$  (Definition 12). By Proposition 7, the  $T$ -transform of  $\mathcal{C}(\mathbf{K}_{\text{At}})$  yields the implicit knowledge-based model  $T(\mathcal{C}(\mathbf{K}_{\text{At}})) = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$ . Use Definition 8 to derive  $\Pi_i^*$  for every  $i \in I$ . By Lemmata 7 to 13,  $\Pi_i^*$  is indeed an explicit possibility correspondence for every  $i \in I$  and  $\overline{T(\mathcal{C}(\mathbf{K}_{\text{At}}))} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$  is a complemented implicit knowledge-based HMS model. Erasing the awareness functions  $(\alpha_i)_{i \in I}$  yields an complemented HMS model  $\langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\Pi_i^*)_{i \in I}, v \rangle$  by Corollary 1. This is the HMS-transform  $HMS(\mathbf{K}_{\text{At}})$ .  $\square$

## 6.2 From Complemented HMS Models to FH Models

To transform a complemented HMS model into a FH model we simply need to consider the upmost space of the lattice of spaces of the HMS model, copy the domain, define

accessibility relations from implicit possibility correspondences as well as the valuation function, and, for every state  $\omega \in S_{\text{At}}$ , construct the awareness set at  $\omega$  by collecting all the formulas that contain the atoms defined in the space where  $\Pi_i(\omega)$  lies.

**Definition 16 (FH-Transform)** *For any complemented HMS model  $\bar{M} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i)_{i \in I}, (\Pi_i)_{i \in I}, v \rangle$ , the FH-transform  $FH(\bar{M}) = \langle I, W_{\text{At}}, (R_{\text{At},i})_{i \in I}, (\mathcal{A}_{\text{At},i})_{i \in I}, V_{\text{At}} \rangle$  is defined by*

- $W_{\text{At}} = S_{\text{At}}$ ,
- $R_{\text{At},i} \subseteq W_{\text{At}} \times W_{\text{At}}$  is such that  $(\omega, \omega') \in R_{\text{At},i}$  if and only if  $\omega' \in \Lambda_i(\omega)$ , for all  $i \in I$ ,
- $\mathcal{A}_{\text{At},i} : W_{\text{At}} \longrightarrow 2^{\mathcal{L}_{\text{At}}}$  is such that  $\mathcal{A}_{\text{At},i}(\omega) = \mathcal{L}_\Phi$  for  $\Phi \subseteq \text{At}$  with  $\Pi_i(\omega) \subseteq S_\Phi$ , for all  $i \in I$ ,
- $V_{\text{At}} : \text{At} \longrightarrow 2^{W_{\text{At}}}$  is such that  $V_{\text{At}}(p) = v(p) \cap S_{\text{At}}$ , for every  $p \in \text{At}$ .

The FH-transform indeed transforms any complemented HMS model into a FH model.

**Proposition 8** *For every complemented HMS model  $\bar{M}$ , the FH-transform  $FH(\bar{M})$  is a FH model for At.*

PROOF. The partitional properties of  $R_{\text{At},i}$  follow in a straightforward way from the partitional properties of  $\Lambda_i$  on  $S_{\text{At}}$ , for every  $i \in I$ .

The fact that Awareness is Generated by Primitive Propositions follows directly from the clause of the FH-transform pertaining to the awareness correspondence  $\mathcal{A}_{\text{At},i}$ . What is left to show is that Agents Know What They Are Aware of. The complemented HMS model satisfies Explicit Measurability. This means in particular that for all  $\omega \in S_{\text{At}}$ ,  $\omega' \in \Lambda_i(\omega)$ , we have  $\Pi_i(\omega') = \Pi_i(\omega)$ . By the clause of the FH-transform pertaining to  $R_{\text{At},i}$ , it follows immediately that  $\omega' \in \Lambda_i(\omega)$  implies  $(\omega, \omega') \in R_{\text{At},i}$ . By the clause of the FH-transform pertaining to the awareness correspondence  $\mathcal{A}_{\text{At},i}$ , we have now that  $\Pi_i(\omega') = \Pi_i(\omega)$  implies  $\mathcal{A}_{\text{At},i}(\omega') = \mathcal{A}_{\text{At},i}(\omega)$ .  $\square$

### 6.3 Equivalence of Complemented HMS and FH Models

Before we can prove an equivalence of HMS and FH models, we need to introduce the semantics of complemented HMS models.

**Definition 17 (Semantics of HMS Models)** *Let  $\bar{M} = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i)_{i \in I}, (\Pi_i)_{i \in I}, v \rangle$  be a complemented HMS model and let  $\omega \in \Omega$ . Satisfaction of  $\mathcal{L}_{\text{At}}$  formulas in  $\bar{M}$  is given by  $\bar{M}, \omega \models \top$  for all  $\omega \in \Omega$  and*

$$\begin{array}{llll}
\bar{M}, \omega \models p & \text{iff} & \omega \in v(p); & \bar{M}, \omega \models a_i \varphi & \text{iff} & S_{\Pi_i(\omega)} \succeq S([\varphi]); \\
\bar{M}, \omega \models \neg \varphi & \text{iff} & \omega \in \neg[\varphi]; & \bar{M}, \omega \models \ell_i \varphi & \text{iff} & \Lambda_i(\omega) \subseteq [\varphi]; \\
\bar{M}, \omega \models \varphi \wedge \psi & \text{iff} & \omega \in [\varphi] \cap [\psi]; & \bar{M}, \omega \models k_i \varphi & \text{iff} & \Pi_i(\omega) \subseteq [\varphi];
\end{array}$$

where  $[\varphi] := \{\omega' \in \Omega : \overline{\mathbf{M}}, \omega' \models \varphi\}$  for all  $\varphi \in \mathcal{L}_{\text{At}}$ .

A couple of comments are in order: First, in HMS models, formulas may have undefined truth value since a formula may not be defined in every state. The same happens in FH models of a category of FH models. For instance, the truth value of  $p$  is not defined for all FH models  $\mathbf{K}_\Phi$  with  $\Phi \not\supseteq p$ . We will return to this issue later in Section 7 when we prove soundness and completeness. Second, recall that for all  $p \in \text{At}$ ,  $v(p)$  is an event, so  $[p]$  is an event in  $\Sigma$ . Negation and intersection of events are events. By Lemmata 1 and 6, explicit knowledge, awareness, and implicit knowledge of events are also events, respectively. Thus, for every  $\varphi \in \mathcal{L}_{\text{At}}$ ,  $[\varphi]$  is an event.

Proposition 4 shows that in complemented HMS models,  $K_i(E) = L_i(E) \cap A_i(E)$ , for any event  $E \in \Sigma$ , so the semantics of  $\mathcal{L}_{\text{At}}$  provided above immediately implies that:

**Proposition 9** *For any complemented HMS model  $\overline{\mathbf{M}}$ ,  $\omega \in \Omega$ ,  $\varphi \in \mathcal{L}_{\text{At}}$ , and  $\Psi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Psi$ ,*

$$\overline{\mathbf{M}}, \omega_\Psi \models k_i \varphi \leftrightarrow (\ell_i \varphi \wedge a_i \varphi).$$

An FH model and its HMS-transform satisfy the same formulas in the language  $\mathcal{L}_{\text{At}}$  with implicit knowledge, explicit knowledge, and awareness as long as these formulas are defined at the corresponding states of the HMS-transform.

**Proposition 10** *For any FH model  $\mathbf{K}_{\text{At}}$  and its HMS-transform  $\text{HMS}(\mathbf{K}_{\text{At}})$ , for all  $w \in W_{\text{At}}$ ,  $\varphi \in \mathcal{L}_{\text{At}}$ , and  $\Phi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Phi$ ,*

$$\mathbf{K}_{\text{At}}, w \Vdash \varphi \text{ if and only if } \text{HMS}(\mathbf{K}_{\text{At}}), w_\Phi \models \varphi.$$

**PROOF.** The proof proceeds by induction on the complexity of the formulas in  $\mathcal{L}_{\text{At}}$ . The case  $\top$  is trivial.

**Base Case:** Consider now any  $p \in \text{At}$ . We have that  $\mathbf{K}_{\text{At}}, w \Vdash p$  if and only if  $w \in V_{\text{At}}(p)$  if and only if (by the Atomic harmony clause of bounded morphisms)  $f_\Phi^{\text{At}}(w) \in V_\Phi(p)$  for all  $\Phi \ni p$  if and only if  $w_\Phi \in v(p)$  for all  $\Phi \ni p$  (by the  $T$ -transform) if and only if  $\text{HMS}(\mathbf{K}_{\text{At}}), w_\Phi \models p$  for all  $\Phi \ni p$ .

**Inductive Step:** The cases  $\neg\varphi$  and  $\varphi \wedge \psi$  are now straightforward. The requirement is now for all  $\Phi \supseteq \text{At}(\varphi)$  and  $\Phi \supseteq \text{At}(\varphi \wedge \psi)$ , respectively.

Now consider the case  $\ell_i \varphi$ : We have  $\mathbf{K}_{\text{At}}, w \Vdash \ell_i \varphi$  if and only if  $\mathbf{K}_{\text{At}}, t \Vdash \varphi$  for all  $t \in W_{\text{At}}$  with  $(w, t) \in R_{\text{At}, i}$  if and only if (by the properties of bounded morphisms)  $\mathbf{K}_\Phi, f_\Phi^{\text{At}}(t) \Vdash \varphi$  for all  $f_\Phi^{\text{At}}(t) \in W_\Phi$  with  $(f_\Phi^{\text{At}}(w), f_\Phi^{\text{At}}(t)) \in R_{\Phi, i}$  for all  $\Phi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Phi$  if and only if (by the induction hypothesis and the  $T$ -transform)  $\text{HMS}(\mathbf{K}_\Phi), t_\Phi \models \varphi$  for all  $t_\Phi \in W_\Phi$  with  $t_\Phi \in \Lambda_i(w_\Phi)$  for all  $\Phi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Phi$  if and only if  $\text{HMS}(\mathbf{K}_\Phi), t_\Phi \models \ell_i \varphi$ .

Next, consider the case  $a_i \varphi$ : We have  $\mathbf{K}_{\text{At}}, w \Vdash a_i \varphi$  if and only if  $\varphi \in \mathcal{A}_{\text{At}, i}(w)$  if and only if (by Awareness consistency of bounded morphisms)  $\varphi \in \mathcal{A}_{\text{At}, i}(w) \cap \mathcal{L}_\Phi = \mathcal{A}_{\Phi, i}(f_\Phi^{\text{At}}(w))$  for all  $\Phi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Phi$  if and only if (by the  $T$ -transform)  $\alpha_i(w_\Phi) \succeq S([\varphi])$  for all  $\Phi \ni \varphi$  if and only if  $\text{HMS}(\mathbf{K}_{\text{At}}), w_\Phi \models a_i \varphi$  for all  $\Phi \subseteq \text{At}$  with  $\text{At}(\varphi) \subseteq \Phi$ .

Finally, the case  $k_i\varphi$  follows now by the induction hypothesis,  $k_i\varphi \leftrightarrow (\ell_i\varphi \wedge a_i\varphi)$ , and Proposition 9.  $\square$

Conversely, we now show that any complemented HMS model and its FH-transform satisfy the same formulas from the language  $\mathcal{L}_{\text{At}}$  with implicit knowledge, explicit knowledge, and awareness.

**Proposition 11** *For any complemented HMS model  $\overline{\mathbf{M}}$  and its FH-transform  $FH(\overline{\mathbf{M}})$ , for all  $\varphi \in \mathcal{L}_{\text{At}}$  and all  $\omega \in S_{\text{At}}$ ,  $\overline{\mathbf{M}}, \omega \models \varphi$  if and only if  $FH(\overline{\mathbf{M}}), \omega \Vdash \varphi$ .*

PROOF. The proof proceeds by straightforward induction on the complexity of formulas in  $\mathcal{L}_{\text{At}}$  starting with the trivial case  $\top$ , the base case  $p \in \text{At}$ , followed by the inductive step for  $\neg\varphi$ ,  $\varphi \wedge \psi$ ,  $\ell_i\varphi$ , and  $a_i\varphi$ . Note that the case  $k_i\varphi$  follows now by Proposition 9 from the preceding steps.  $\square$

## 6.4 From FH to Implicit Knowledge-based HMS Models

We now focus on the relationship between implicit knowledge-based HMS and FH models. This relationship is even simpler than for completed HMS and FH models since implicit knowledge-based HMS models are arguably already closer to FH models than complemented HMS models. This is due to taking implicit knowledge and the awareness functions as primitives.

We define a version of HMS transformation that is “truncated” after the  $T$ -transformation. It just keeps the first two steps of the HMS transformation.

**Definition 18 (Truncated HMS-Transform)** *For any FH model  $\mathbf{K}_{\text{At}}$ , the truncated HMS-transform  $HMS^*(\mathbf{K}_{\text{At}}) = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i^*)_{i \in I}, (\alpha_i^*)_{i \in I}, v \rangle$  is defined by the first two steps of the HMS-transform (cf. Definition 15).*

From Proposition 7 follows now immediately:

**Corollary 3** *For any FH model  $\mathbf{K}_{\text{At}}$ , the truncated HMS-transform  $HMS^*(\mathbf{K}_{\text{At}})$  is an implicit knowledge-based HMS model.*

## 6.5 From Implicit Knowledge-based HMS to FH Models

**Definition 19 (FH\*-Transform)** *For any implicit knowledge-based HMS model  $\overline{\mathbf{M}}^* = \langle I, \{S_\Phi\}_{\Phi \subseteq \text{At}}, (r_\Psi^\Phi)_{\Psi \subseteq \Phi \subseteq \text{At}}, (\Lambda_i)_{i \in I}, (\alpha_i)_{i \in I}, v \rangle$ , the FH\*-transform  $FH^*(\overline{\mathbf{M}}^*) = \langle I, W_{\text{At}}, (R_{\text{At},i})_{i \in I}, (\mathcal{A}_{\text{At},i})_{i \in I}, V_{\text{At}} \rangle$  is defined like the FH-transform except that the clause for the awareness correspondence is replaced by, for any  $i \in I$ ,*

- $\mathcal{A}_{\text{At},i} : W_{\text{At}} \longrightarrow 2^{\mathcal{L}_{\text{At}}}$  is such that  $\mathcal{A}_{\text{At},i}(\omega) = \mathcal{L}_\Phi$  for  $\Phi \subseteq \text{At}$  with  $\alpha_i(\omega) = S_\Phi$ .

The  $\text{FH}^*$ -transform indeed transforms any implicit knowledge-based HMS model into a FH model.

**Proposition 12** *For every implicit knowledge-based HMS model  $\overline{\mathbf{M}}^*$ , the  $\text{FH}^*$ -transform  $\text{FH}^*(\overline{\mathbf{M}}^*)$  is a FH model for At.*

PROOF. The proof is analogous to the proof of Proposition 8 except we need to modify the clause relating to the awareness correspondence.

The fact that Awareness is Generated by Primitive Propositions follows directly from the clause of the  $\text{FH}^*$ -transform pertaining the awareness correspondence  $\mathcal{A}_{\text{At},i}$ . What is left to show is that Agents Know What They Are Aware of. The implicit knowledge-based HMS model satisfies Awareness Measurability. This means in particular that for all  $\omega \in S_{\text{At}}$ ,  $\omega' \in \Lambda_i(\omega)$ , we have  $\alpha_i(\omega') = \alpha_i(\omega)$ . By the clause of the  $\text{FH}^*$ -transform pertaining to  $R_{\text{At},i}$ , it follows immediately that  $\omega' \in \Lambda_i(\omega)$  implies  $(\omega, \omega') \in R_{\text{At},i}$ . By the clause of the  $\text{FH}^*$ -transform pertaining the awareness correspondence  $\mathcal{A}_{\text{At},i}$ , we have now that  $\alpha_i(\omega') = \alpha_i(\omega)$  implies  $\mathcal{A}_{\text{At},i}(\omega') = \mathcal{A}_{\text{At},i}(\omega)$ .  $\square$

## 6.6 Equivalence of Implicit Knowledge-based HMS and FH Models

To prove the equivalence, we need to introduce the semantics of  $\mathcal{L}_{\text{At}}$  over implicit knowledge-based HMS models. The semantics is identical to the semantics over complemented HMS models, except for the awareness clause.<sup>37</sup>

**Definition 20** *Satisfaction of  $\mathcal{L}_{\text{At}}$  formulas in an implicit knowledge-based HMS model  $\mathbf{M}^*$  is given like for complemented HMS models except that we have  $\mathbf{M}^*, \omega \models a_i\varphi$  if and only if  $\alpha_i(\omega) \succeq S([\varphi])$ .*

From this semantics and the syntactic definition  $k_i\varphi := \ell_i\varphi \wedge a_i\varphi$ , it follows that  $\mathbf{M}^*, \omega \models k_i\varphi$  if and only if  $\mathbf{M}^*, \omega \models \ell_i\varphi$  and  $\mathbf{M}^*, \omega \models a_i\varphi$ . The comments about the semantics of complemented HMS models from Section 6.3 (right below the semantics definition) as well as Proposition 9, also hold for implicit knowledge-based HMS models.

An FH model and its truncated HMS-transform satisfy the same formulas in the language  $\mathcal{L}_{\text{At}}$  with implicit knowledge, explicit knowledge, and awareness with the provision that these formulas are defined at the corresponding states of the implicit knowledge-based HMS-transform. This follows directly from the proof of Proposition 10.<sup>38</sup>

<sup>37</sup>We slightly abuse notation and use the same symbol for the satisfaction relation as we did for complemented HMS models.

<sup>38</sup>The proof of Proposition 10 makes use of the HMS-transform, which in turn makes use of the  $T$ -transform. When arguments in that proof are stopped after making use of the  $T$ -transform, the proof of the corollary is completed.

**Corollary 4** *For any FH model  $K_{At}$  and its HMS-transform  $HMS(K_{At})$ , for all  $w \in W_{At}$ ,  $\varphi \in \mathcal{L}_{At}$ , and  $\Phi \subseteq At$  with  $At(\varphi) \subseteq \Phi$ ,*

$$K_{At}, w \Vdash \varphi \text{ if and only if } HMS(K_{At}), w_\Phi \models \varphi.$$

Any implicit knowledge-based HMS model and its FH\*-transform satisfy the same formulas from the language  $\mathcal{L}_{At}$  with implicit knowledge, explicit knowledge, and awareness.

**Proposition 13** *For any implicit knowledge-based HMS model  $M^*$  and its FH\*-transform  $FH^*(M^*)$ , for all  $\varphi \in \mathcal{L}_{At}$  and all  $\omega \in S_{At}$ ,*

$$M^*, \omega \models \varphi \text{ if and only if } FH^*(M^*), \omega \Vdash \varphi.$$

PROOF. The proof proceeds by straightforward induction on the complexity of formulas in  $\mathcal{L}_{At}$  starting with the trivial case  $\top$ , the base case  $p \in At$ , followed by the inductive step for  $\neg\varphi$ ,  $\varphi \wedge \psi$ ,  $\ell_i\varphi$ ,  $a_i\varphi$ , and  $k_i\varphi$ .  $\square$

## 7 Logic of Propositional Awareness

In this section, we explore the implications of the prior sections for axiomatizations of both the category of FH models and HMS models (both complemented and implicit knowledge-based). In particular, we show that the Logic of Propositional Awareness is sound and complete with respect to the class of complemented HMS models. This is the first axiomatization of HMS models that features also the notion of implicit knowledge. Previous axiomatizations of HMS models (Heifetz, Meier, and Schipper, 2008, Halpern and Rêgo, 2008) were confined to explicit knowledge and awareness only as the model did not contain a notion of implicit knowledge. We also show that the Logic of Propositional Awareness is sound and complete with respect to the class of implicit knowledge-based HMS models. This is the first axiomatization of implicit knowledge-based HMS models. Finally, it is also sound and complete with respect to the class of categories of FH models.

**Definition 21** *The logic LPA is the smallest set of  $\mathcal{L}_{At}$  formulas that contains the axioms in, and is closed under the inference rules of, Table 1.*

In Table 1, the Explicit Knowledge axiom captures the definition of explicit knowledge proposed by Fagin and Halpern (1988). Axioms 1-5 capture a propositionally generated notion of awareness, while axioms 11-12 characterize individuals know what they are aware of (the numbering of awareness axioms is taken from Halpern, 2001). Axioms K, T, 4, and 5 constitute the standard notion of S5 knowledge (see e.g., Rendsvig and Symons, 2021). Lastly, Table 1 contains inference rules and substitution that are standard in modal logic.

Recall that in a Kripke model or FH model, a formula is valid if it is true in every state. However, a formula  $\varphi \in \mathcal{L}_{At}$  is not even defined at states of the FH model  $K_\Psi$  with

Table 1: Axioms and inference rules of the Logic of Propositional Awareness (LPA), for a propositionally determined notion of awareness and axioms for an  $S5$  logic.

All substitution instances of propositional logic, including the formula $\top$	
$(\ell_i\varphi \wedge (\ell_i\varphi \rightarrow \ell_i\psi)) \rightarrow \ell_i\psi$	(K, Distribution)
$k_i\varphi \leftrightarrow (\ell_i\varphi \wedge a_i\varphi)$	(Explicit Knowledge)
$a_i(\varphi \wedge \psi) \leftrightarrow (a_i\varphi \wedge a_i\psi)$	(A1, Awareness Distribution)
$a_i\neg\varphi \leftrightarrow a_i\varphi$	(A2, Symmetry)
$a_ik_j\varphi \leftrightarrow a_i\varphi$	(A3, Awareness of Explicit Knowledge)
$a_ia_j\varphi \leftrightarrow a_i\varphi$	(A4, Awareness Reflection)
$a_ill_j\varphi \leftrightarrow a_i\varphi$	(A5, Awareness of Implicit Knowledge)
$a_i\varphi \rightarrow \ell_ia_i\varphi$	(A11, Awareness Introspection)
$\neg a_i\varphi \rightarrow \ell_i\neg a_i\varphi$	(A12, Unawareness Introspection)
From $\varphi$ and $\varphi \rightarrow \psi$ , infer $\psi$	(Modus Ponens)
From $\varphi$ infer $\ell_i\varphi$	(K-Inference)
$\ell_i\varphi \rightarrow \varphi$	(T, Truth)
$\ell_i\varphi \rightarrow \ell_i\ell_i\varphi$	(4, Positive Introspection)
$\neg\ell_i\varphi \rightarrow \ell_i\neg\ell_i\varphi$	(5, Negative Introspection)

$\text{At}(\varphi) \not\subseteq \Psi$ . Similarly, as we remarked earlier when introducing the semantics for HMS models, a formula may not be defined in every state of a HMS model. We say that  $\varphi$  is defined in state  $\omega$  in the complemented HMS model  $\overline{\mathcal{M}}$  if  $\omega \in \bigcap_{p \in \text{At}(\varphi)} (v(p) \cup \neg v(p))$  (and analogously for implicit knowledge-based HMS models). Similarly, we say that  $\varphi$  is defined in the FH model  $\mathcal{K}_\Psi$  if  $\text{At}(\varphi) \subseteq \Psi$ .

Now we say that a formula  $\varphi$  is valid in the complemented HMS model  $\overline{\mathcal{M}}$  if  $\overline{\mathcal{M}}, \omega \models \varphi$  for all  $\omega$  in which  $\varphi$  is defined (and analogously for the implicit knowledge-based HMS model). Similarly, we say that  $\varphi$  is valid in the category of FH models  $\mathcal{C}(\mathcal{K}_\Psi)$  if  $\mathcal{K}_\Psi, w \models \varphi$  for all  $w \in W_\Psi$  for all  $\mathcal{K}_\Psi$  in  $\mathcal{C}(\mathcal{K}_{\text{At}})$  for which  $\varphi$  is defined. A formula is valid in a class of complemented HMS models  $\mathcal{M}$  if it is valid in every complemented HMS model of the class (and analogously for the class of implicit knowledge-based HMS models). A formula is valid in a class of categories of FH models  $\mathfrak{C}$  if it is valid in every category of FH models of the class.

A proof in an axiom system consists of a sequence of formulas, where each formula in the sequence is either an axiom in the axiom system or follows from the prior formula in the sequence by an application of an inference rule of the axiom system. A proof of a formula  $\varphi$  is a proof where the last formula of the sequence is  $\varphi$ . A formula  $\varphi$  is provable in an axiom system, if there is a proof of  $\varphi$  in the axiom system. An axiom system is sound for the language  $\mathcal{L}_{\text{At}}$  with respect to a class of complemented HMS models  $\mathcal{M}$  if every formula in  $\mathcal{L}_{\text{At}}$  that is provable in the axiom system is valid in every complemented HMS model of the class  $\mathcal{M}$  (and analogously for the class of implicit knowledge-based HMS models). Similarly, an axiom system is sound for the language  $\mathcal{L}_{\text{At}}$  with respect to a class of categories of FH models  $\mathfrak{C}$  if every formula in  $\mathcal{L}_{\text{At}}$  that is provable in the axiom system is valid in every category of FH models of the class  $\mathfrak{C}$ . An axiom system is complete for the language  $\mathcal{L}_{\text{At}}$  with respect to a class of complemented HMS models  $\mathcal{M}$  if every formula in  $\mathcal{L}_{\text{At}}$  that is valid in  $\mathcal{M}$  is provable in the axiom system (and analogously



for the class of implicit knowledge-based HMS models). Similarly, an axiom system is complete for the language  $\mathcal{L}_{At}$  with respect to a class of categories of FH models  $\mathfrak{C}$  if every formula in  $\mathcal{L}_{At}$  that is valid in  $\mathfrak{C}$  is provable in the axiom system.

**Corollary 5** *LPA is sound and complete with respect to*

1. *the class of categories of FH models,*
2. *the class of complemented HMS models,*
3. *the class of implicit knowledge-based HMS models.*

Fagin and Halpern (1988), Halpern (2001), and Halpern and Rêgo (2008) claim that LPA is sound and complete with respect to the class of FH models. The proof of (1.) now follows from invariance of modal satisfaction relative to sublanguages between FH models in each category of FH models, i.e., Proposition 6. The proof of (2.) follows from Propositions 10 and 11. The proof of (3.) follows from Corollary 4 and Proposition 13.

## 8 Discussion

We enriched HMS models with a notion of implicit knowledge. We showed how implicit knowledge can be defined in HMS models so that it is consistent with explicit knowledge. We also introduced a variant of HMS models based on implicit knowledge and awareness and showed how to derive explicit knowledge. This demonstrates that explicit knowledge and implicit knowledge are in some sense interdefinable in HMS models. By introducing implicit knowledge into HMS models, we arguably made them “closer” to FH models. At the same time, we made FH models “closer” to HMS models by forming a category of FH models that differ by the language with bounded morphisms between them. This allowed us to consider subjective FH models, i.e., models that agents themselves could use at particular states to analyze their situation.

The constructions also allowed us to consider the relation between FH models and HMS models, not just with respect to explicit knowledge and awareness, as in the prior literature, but also with respect to implicit knowledge. We showed an equivalence between FH and HMS models by transforming a model into the other and vice versa, and by showing that each model and its transform satisfy the same formulas. This equivalence is used to show that the Logic of Propositional Awareness is sound and complete with respect to the class of HMS models. Compared to the prior literature, this axiomatization is now for a language that also features implicit knowledge.

The relations between various models of awareness in the literature are depicted in Figure 2. Beside FH models of Fagin and Halpern (1988) and HMS models of Heifetz, Meier, and Schipper (2006, 2008), we consider generalized standard models by Modica and Rustichini (1999), information structures with unawareness by Li (2009), object-based unawareness models by Board and Chung (2021), and Kripke lattices by Belardinelli and Rendsvig (2022). Equivalences hold for various languages also shown in the

Figure 2: Relations between Approaches to Awareness

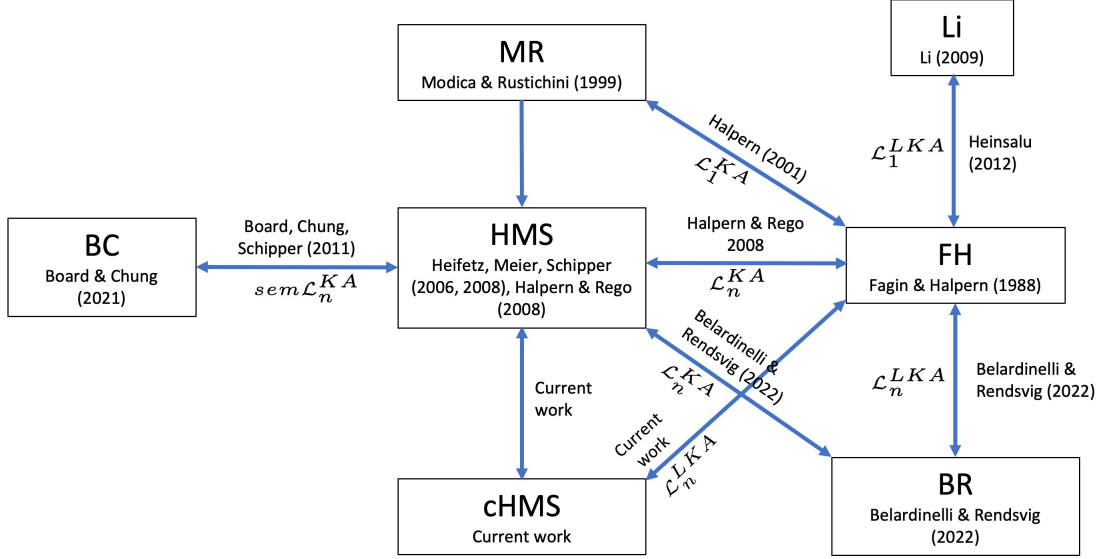


figure. We indicate the implicit, explicit, and awareness modality by superscripts  $L$ ,  $K$ , and  $A$ , respectively. Some structures like Modica and Rustichini (1999) and Li (2009) feature just a single agent. We indicate this with the subscript "1" for single agent and "n" for multiple agents. For instance,  $\mathcal{L}_n^{L,K,A}$  is the language featuring multiple agents, implicit knowledge, explicit knowledge, and awareness. The equivalence between Board and Chung (2021) is shown only at the level of semantics, i.e., at the level of events. The relation between Modica and Rustichini (1999) and Heifetz, Meier, and Schipper (2008) indicates that latter axiomatization can be seen as a multi-agent version of former. All shown relations pertain to rich structures featuring partitional knowledge and awareness generated by primitive propositions.

Recently, Schipper (2022) extended HMS models to awareness of unawareness by introducing quantified events. It would be straightforward to complement his structure with implicit knowledge as defined in the current work. Agents could then reason about the existence of their own implicit knowledge that they are not aware of. Such reasoning appears to be similar to the notion of speculative knowledge in van Ditmarsch et al. (2018). Awareness-of-unawareness structures with implicit knowledge would also allow for a better comparison to awareness structures with quantification of formulas for modeling reasoning about knowledge of unawareness (Halpern and Rêgo, 2009, 2012), object-based unawareness (Board and Chung, 2021), and quantified neighborhood structures with awareness (Sillari, 2008), all of which feature notions of implicit knowledge.

We view our work as a first step towards a more comprehensive study of notions of implicit knowledge outlined in the dialogue of the introduction. For instance, to what extent can implicit knowledge satisfies less stringent properties than explicit knowledge (or vice versa) while still being coherent with each other? Could such a notion of implicit knowledge be useful to think about implicit biases (Brownstein, 2019) or implicit cogni-

tion in cognitive psychology (Augusto, 2010)? Second, is there a notion of explicability, a kind of second layer of awareness, that could serve to define a notion of tacit knowledge (Polanyi, 1962)? These questions guide our current work.

## References

- [1] Augusto, L. M. (2010). Unconscious knowledge: A survey, *Advances in Cognitive Psychology* 6, 116–141.
- [2] Aumann, R. (1999). Interactive epistemology I: Knowledge, *International Journal of Game Theory* 28, 263–300.
- [3] Auster, S. (2013). Asymmetric awareness and moral hazard, *Games and Economic Behavior* 82, 503–521.
- [4] Auster, S., Pavoni, N. (2021a). Optimal delegation and information transmission under limited awareness, University of Bonn & Bocconi University.
- [5] Auster, S., Pavoni, N. (2021b). Limited awareness and financial intermediation, University of Bonn & Bocconi University.
- [6] Bargh, J.A., Chen, M., Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action, *Journal of Personality and Social Psychology* 71, 230–244.
- [7] Belardinelli, G., Rendsvig K.R. (2022) Awareness logic: Kripke lattices as a middle ground between syntactic and semantic models. *Journal of Logic and Computation*, exac009.
- [8] Blackburn, P., de Rijke, M., Venema, Y. (2001). *Modal Logic*. Cambridge University Press.
- [9] Board, O., Chung, K.S. (2021). Object-based unawareness: Axioms, *Journal of Mechanism and Institutional Design* 6, 1–36.
- [10] Board, O., Chung K.S., Schipper, B.C. (2011). Two models of unawareness: Comparing the object-based and subjective-state-space approaches, *Synthese* 179, 13–34.
- [11] Brownstein, M. (2019). Implicit Bias, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/fall2019/entries/implicit-bias/>
- [12] Brownstein, M., Madva, A., Gawronski, B. (2019). What do implicit measures measure? *WIREs Cogn Sci*.
- [13] Bryan, K, Ryall, M., Schipper, B.C. (2022). Value-capture in the face of known and unknown unknowns, *Strategy Science* 7, 157–189.
- [14] Chung, K., Fortnow, L. (2016). Loopholes, *Economic Journal* 126, 1774–1797.
- [15] De Houwer, J. (2006). What Are Implicit Measures and Why Are We Using Them? In R. W. Wiers & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction*, 11–28.
- [16] Dekel, E., Lipman, B., Rustichini, A. (1998). Standard state-space models preclude unawareness, *Econometrica* 66, 159–173.
- [17] Dominiak, A., Tserenjigmid, G. (2018). Belief consistency and invariant risk preferences, *Journal of Mathematical Economics* 79, 157–162.
- [18] Doyen, S., Klein, O., Pichon, C.-L., Cleeremans, A. (2012). Behavioral priming: It’s all in the mind, but whose mind? *PLoS ONE* 7, e29081.
- [19] Fagin, R., Halpern, J. (1988). Belief, awareness, and limited reasoning, *Artificial Intelligence* 34, 39–76.

- [20] Fagin, R., Halpern, J., Vardi, M. (1986). What can machines know? On the epistemic properties of machines, AAAI 1986.
- [21] Fagin, R., Halpern, J., Moses, Y., Vardi, M. (1995). Reasoning about knowledge, MIT Press.
- [22] Feinberg, Y. (2021). Games with unawareness, B.E. Journal of Theoretical Economics 21, 433–488.
- [23] Filiz-Ozbay, E. (2012). Incorporating unawareness into contract theory, Games and Economic Behavior 76, 181–194.
- [24] Francetich, A., Schipper, B.C. (2021). Rationalizable screening and disclosure under unawareness, University of California, Davis.
- [25] Freud, S. (1915). Das Unbewusste, Internationale Zeitschrift für Ärztliche Psychoanalyse 3, 189–203, 257–269.
- [26] Galanis, S. (2018). Speculation under unawareness, Games and Economic Behavior 109, 598–615.
- [27] Galanis, S. (2013). Unawareness of theorems, Economic Theory 52, 41–73.
- [28] Galanis, S. (2011). Syntactic foundations of unawareness of theorems, Theory and Decision 71, 593–614.
- [29] Gawronski, B., Ledgerwood, A., Eastwick, P. (2022). Implicit bias  $\neq$  bias on implicit measures, Psychological Inquiry 33, 139–155.
- [30] Gawronski, B., Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In: Blanton, H., LaCroix, J.M., Webster, G.D. (Eds.), Measurement in social psychology, 29–55.
- [31] Gigerenzer, G. (2001). The adaptive toolbox, in: Gigerenzer, G., Selten, R. (Eds.). Bounded rationality: The adaptive toolbox, MIT Press, 37–50.
- [32] Guarino, P. (2020). An epistemic analysis of dynamic games with unawareness, Games and Economic Behavior 120, 257–288.
- [33] Grant, S., Quiggin, J. (2013). Inductive reasoning about unawareness, Economic Theory 54, 717–755.
- [34] Grant, S., Kline, J.J., O’Callaghan, P., Quiggin, J. (2015). Sub-models for interactive unawareness, Theory and Decision 79, 601–613.
- [35] Greenwald, A.G., Banaji, M.R., Nosek, B.A. (2015). Statistically small effects of the implicit association test can have societally large effects, Journal of Personality and Social Psychology 108, 553–561.
- [36] Greenwald, A.G., McGhee, D. E., Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. Journal of Personality and Social Psychology 74, 1464–1480.
- [37] Hahn, A., Judd, C.M., Hirsh, H.K., Blair, I.V. (2014). Awareness of implicit attitudes, Journal of Experimental Psychology:General 143, 1369–1392.
- [38] Halpern, J.Y. (1987). Using reasoning about knowledge to analyze distributed systems, Annual Review of Computer Science 2, 37–68.
- [39] Halpern, J.Y. (2001). Alternative Semantics for Unawareness. Games and Economic Behavior 37, 321–339.
- [40] Halpern, J.Y., Rêgo, L.C. (2008). Interactive unawareness revisited. Games and Economic Behavior 62, 232–262.
- [41] Halpern, J.Y., Rêgo, L.C. (2009). Reasoning about knowledge of unawareness, Games and Economic Behavior 67, 503–525.
- [42] Halpern, J.Y., Rêgo, L.C. (2012). Reasoning about knowledge of unawareness revisited, Mathematical Social Sciences 65, 73–84.

- [43] Halpern, J.Y., Rêgo, L.C. (2014). Extensive games with possibly unaware players, *Mathematical Social Sciences* 70, 42–58.
- [44] Heifetz, A., Meier, M., Schipper, B.C. (2006). Interactive unawareness, *Journal of Economic Theory* 130, 78–94.
- [45] Heifetz, A., Meier, M., Schipper, B.C. (2008). A canonical model for interactive unawareness, *Games and Economic Behavior* 62, 305–324.
- [46] Heifetz, A., Meier, M., Schipper, B.C. (2013a). Unawareness, beliefs, and speculative trade, *Games and Economic Behavior* 77, 100–121.
- [47] Heifetz, A., Meier, M., Schipper, B.C. (2013b). Dynamic unawareness and rationalizable behavior, *Games and Economic Behavior* 81, 50–68.
- [48] Heifetz, A., Meier, M., Schipper, B.C. (2021). Prudent rationalizability in generalized extensive-form games with unawareness, *B.E. Journal of Theoretical Economics* 21, 525–556.
- [49] Heinsalu, S. (2012). Equivalence of the information structure with unawareness to the logic of awareness, *Journal of Economic Theory* 147, 2453–2468.
- [50] Hintikka, J. (1975). Impossible possible worlds vindicated, *Journal of Philosophical Logic* 4, 475–484.
- [51] Karni, E., Vierø, M.-L. (2013). “Reverse Bayesianism”: A choice-based theory of growing awareness, *American Economic Review* 103, 2790–2810.
- [52] Lakemeyer, G. (1986). Steps towards a first-order logic of explicit and implicit belief, *TARK* 1986.
- [53] Levesque, H.J. (1984). A logic of implicit and explicit belief, *AAAI Proceedings*, 198–202.
- [54] Li J. (2009). Information structures with unawareness, *Journal of Economic Theory* 144, 977–993.
- [55] Lorini, E. (2020). Rethinking epistemic logic with belief bases. *Artificial Intelligence* 282, 203233.
- [56] Meier, M., Schipper, B.C. (2014a). Bayesian games with unawareness and unawareness perfection, *Economic Theory* 56, 219–249.
- [57] Meier, M., Schipper, B.C. (2014b). Speculative trade under unawareness - The infinite case, *Economic Theory Bulletin* 2, 147–160.
- [58] Meier, M., Schipper, B.C. (2023). Conditional dominance in games with unawareness, *University of California, Davis*.
- [59] Meissner, F., Grigutsch, L.A., Koranyi, N., Müller, F., Rothermund, K. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions, *Frontiers in Psychology* 10, Article 2483.
- [60] Mertens, J.-F., Zamir, S. (1984). Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory* 14, 1–29.
- [61] Modica, S., Rustichini A. (1999). Unawareness and partitional information structures, *Games and Economic Behavior* 27, 265–298.
- [62] Moss, L. (1999). Coalgebraic logic, *Annals of Pure and Applied Logic* 96, 277–317.
- [63] Osborne, M., Rubinstein A. (1994). *A course in game theory*, MIT Press.
- [64] Perea, A. (2022). Common belief in rationality in games with unawareness, *Mathematical Social Sciences* 119, 11–30.
- [65] Polanyi, M. (1962). *Personal knowledge*, Routledge, corrected edition.
- [66] Pram, K., Schipper, B.C. (2023). Efficient mechanisms under unawareness, *University of California, Davis*.

- [67] Rêgo, L.C., Halpern, J. (2012). Generalized solution concepts in games with possibly unaware players, *International Journal of Game Theory* 41, 131–155.
- [68] Rêgo, L.C., Vieira, G.I.A. (2020). Interactive unawareness in the graph model for conflict resolution, *IEEE Transactions on Systems, Man, and Cybernetics* 50, 829–839.
- [69] Rendsvig, R., Symons, J. “Epistemic Logic”, *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/logic-epistemic/>.
- [70] Roediger, H., McDermott, K. (1993). Implicit memory in normal human subjects. *Handbook of Neuropsychology*. Amsterdam: Elsevier Science Publishers: 63-131.
- [71] Roth, G. (2003). *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert*, Suhrkamp Verlag.
- [72] Stalnaker, R. (1991). The Problem of Logical Omniscience, I. *Synthese*, 89(3), 425–440.
- [73] Schimmack, U. (2022). Implicit bias  $\neq$  biased in implicit measures, <https://replicationindex.com/2022/11/18/implicitbiasdefinition/>.
- [74] Schimmack, U. (2021). Reevaluating the predictive validity of the race implicit association test, <https://replicationindex.com/2021/06/13/predictive-validity-race-iat/>.
- [75] Schimmack, U., Heene, M., Kesavan, K. (2017). Reconstruction of a train wreck: How priming research went off the rails, <https://replicationindex.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-off-the-rails/>.
- [76] Schipper, B.C. (2022). *Interactive awareness of unawareness*, University of California, Davis.
- [77] Schipper, B.C. (2021). Discovery and equilibrium in games with unawareness, *Journal of Economic Theory* 198, 105365.
- [78] Schipper, B.C. (2016). *Network formation in a society with fragmented knowledge and awareness*, University of California, Davis.
- [79] Schipper, B.C. (2015). Awareness, in: *Handbook of epistemic logic*, Chapter 3, H. van Ditmarsch, J.Y. Halpern, W. van der Hoek and B. Kooi (Eds.), College Publications, London, 2015, 77–146.
- [80] Schipper, B.C. (2014). Preference-based unawareness, *Mathematical Social Sciences* 70, 34–41.
- [81] Schipper, B.C. (2013). Awareness-dependent subjective expected utility, *International Journal of Game Theory* 42, 725–753.
- [82] Schipper, B.C., Woo, H.Y. (2019). Political awareness and microtargeting of voters in electoral competition, *Quarterly Journal of Political Science* 14, 41–88.
- [83] Schipper, B.C., Zhou, H. (2021). *Insiders trading on unknown unknowns*, University of California, Davis.
- [84] Selten, R. (2002). What is bounded rationality? in: Gigerenzer, G., Selten, R. (Eds.). *Bounded rationality. The adaptive toolbox*, MIT Press.
- [85] Sillari, G. (2008). Quantified logic of awareness and impossible possible worlds, *Review of Symbolic Logic* 1, 514–529.
- [86] Simon, H. (1957). *Models of man*, Wiley.
- [87] van Ditmarsch, H., French, T., Velázquez-Quesada, F., Wáng, Y. N. (2018). Implicit, explicit and speculative knowledge, *Artificial Intelligence* 256, 35–67.
- [88] Velázquez-Quesada, F.R. (2014). Dynamic epistemic logic for implicit and explicit beliefs, *Journal of Logic, Language, and Information* 23, 107–140.
- [89] Velázquez-Quesada, F.R. (2013). Explicit and implicit knowledge in neighbourhood models, in: Grossi, D., Roy, O., Huang, H. (Eds.), *LORI 2013*, Springer, 239–252.
- [90] von Thadden, E.-L., Zhao, X.J. (2012). Incentives for unaware agents, *Review of Economic Studies* 79, 1151–1174.