A formal model for Explicit Knowledge as Awareness-of plus Awareness-that

Fernández-Fernández, C.; Velázquez-Quesada, F.R.

Publication date
2019

Document Version
Author accepted manuscript

Published in
The Logica Yearbook 2018

A formal model for Explicit Knowledge as Awareness-of plus Awareness-that*

Claudia Fernández-Fernández†1 and Fernando R. Velázquez-Quesada[2]

[1] *University of Málaga.*, `cffernandez@uma.es`
[2] *Institute for Logic, Language and Computation, Universiteit van Amsterdam*, `F.R.VelazquezQuesada@uva.nl`

Abstract

In the context of the problem of logical omniscience, several frameworks have been proposed to model the knowledge of 'real' agents with limited reasoning abilities. One of the most important, *awareness logic*, relies on the concept of *awareness* for distinguishing what the agent 'truly' knows from what she could get if she were aware of all formulas. Still, the notion of awareness can be interpreted in different ways: it can be understood as what the agent simply entertains, without having any attitude in favour or against (*awareness of*), but also as what she has consciously recognised as true (*awareness that*). This paper proposes a formal framework that captures these two interpretations of the notion of awareness, discussing the further epistemic notions that arise from their combination (e.g., *implicit knowledge* and *explicit knowledge*) while also studying their properties and the way they interact with one another.
Keywords: Awareness, Explicit Knowledge, Awareness Logic, Neighbourhood Semantics, Dynamic Epistemic Logic.

1 Introduction

Since the problem of logical omniscience was identified (Hintikka 1962; see also Stalnaker 1991), several frameworks have been proposed to model the knowledge of 'real' agents with limited reasoning abilities. To do so, knowledge has been typically split into *explicit* and *implicit*, with the former being the 'real' knowledge the agent has, and the latter being the knowledge an ideal agent would obtain.

Among the many different proposals, the *awareness logic* of Fagin and Halpern (1988) has been one of the most successful. It relies on the insight that, for an agent to know that certain φ indeed holds, it is not enough for φ to

*Published as Fernández-Fernández and Velázquez-Quesada (2019).
†Partially supported by the Spanish Ministry of Science Project number TIN15-70266-C2-P-1 and the European Reginal Fund Development (ERFD).

be the case in all her epistemic alternatives (as the standard epistemic logic of Hintikka 1962 requires): the agent should be also *aware of* it. This simple but powerful idea has proven to be useful in Philosophy, Computer Science and Economics (see, e.g., the comprehensive handbook chapter Schipper 2015).

Still, the notion of awareness can be interpreted in different ways (cf. Dretske 1993): it can be understood as what the agent simply entertains, without having any attitude in favour or against (*awareness-of*), but also as what she has consciously recognised as true (*awareness-that*). Thus, lacking of awareness can have different meanings. On the one hand, potential lacking of *awareness-of* yields agents who, while possibly not aware of all involved possibilities, are still ideal reasoners within the realm of what they entertain (see, e.g., the original Fagin and Halpern 1988, and also Heifetz et al. 2008, Halpern and Rêgo 2008). On the other hand, potential lacking of *awareness-that* yields agents which entertain all relevant possibilities, and still might not have been able to realise that a certain ψ is the case despite knowing (explicitly) both φ and $\varphi \rightarrow \psi$ (see, e.g., Konolige 1984b, Velázquez-Quesada 2013).

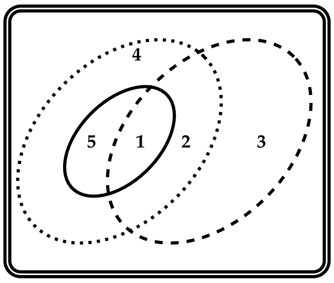
Fernández-Fernández and Velázquez-Quesada (2019) proposed, at an intuitive level, a setting understanding explicit knowledge as the combination of both *awareness-of* and *awareness-that*: in order to know a given φ explicitly, the agent needs to entertain the possibility of φ, but she also needs to recognise that the formula is indeed the case.[1] [2] This is a crucial feature, as considering these two kinds of awareness allows us to separate the mere fact of entertaining some information (being *aware of* φ; just a matter of attention) from recognising (acknowledging/accepting) that some φ is indeed the case (being *aware that* φ).

A setting considering both *awareness-of* and *awareness-that* gives rise to further epistemic concepts, and the diagram on Figure 1 provides a visual aid for this. While the small ellipse near the centre contains what the agent is aware that, the large dashed ellipse on the right contains what the agent is aware of. Two further 'big areas' arise: the logical consequences of what the agent is aware that (the large dotted ellipse on the left), and all truthful information the agent might become aware of (the whole domain). The former can be seen as the agent's implicit knowledge under acts of deductive inference (what she would be aware-that if she were to perform all possible deductive inferences); the latter can be seen as the agent's implicit knowledge under acts of becoming aware (what she would entertain if she became aware of all relevant possibilities). Together, they define the regions **1** to **5**, described by the text next to the diagram.

The present proposal provides a formal logical framework capturing these intuitive notions. First, it provides a model in which *awareness that* is depicted by means of a neighbourhood function (Scott 1970, Montague 1970; see Pacuit 2017 for a modern presentation), with each world's neighbourhood understood as a list containing the semantic representation of the formulas the agent has acknowledged as true. Then, *awareness of* is depicted by a single set of atomic propositions (cf. Fagin and Halpern 1988), understood as the atoms defining the agent's current language. Explicit knowledge is then defined as what the agent is both aware of and aware that, with implicit knowledge defined as what

[1] Cf. the proposal in Grossi and Velázquez-Quesada (2015).

[2] In line with the epistemological view of knowledge as 'Justified True Belief', asking for the agent to also have a *justification* supporting her explicit knowledge.



1 - What the agent is aware that and aware of (i.e., what she explicitly knows).

2 - What the agent is entertaining, has not recognised as true, but will after deductive reasoning.

3 - What the agent is entertaining, has not recognised as true, and is outside the scope of deductive reasoning.

4 - What the agent could deduce if she became aware of it.

5 - What the agent has recognised as true but is not currently entertaining.

Figure 1: Combining *awareness of* and *awareness that*.

the agent would know explicitly after performing every possible deductive inference. The proposal also provides a formal language for describing such structures, allowing a formal discussion of the concepts' properties and subtle interactions. The main aim is not only to provide a slightly different approach for modelling the knowledge of more 'real' agents, but also to shed light to other epistemic concepts that arise when combining these two types of awareness.

2 The formal framework

Let $\mathtt{P}$ be a non-empty enumerable set of atomic propositions.

Definition 1 (Awareness neighbourhood model) An *awareness neighbourhood model* (*ANM*) is a tuple $M = \langle W, N, V, A \rangle$ where ***(i)*** W, sometimes denoted as $\mathcal{D}_M$, is a non-empty set (whose elements are called *possible worlds*), ***(ii)*** $N : W \to \wp(\wp(W))$ is a *neighbourhood function* (assigning a set of sets of worlds to each possible world, with $N(w)$ called the *neighbourhood* of w), ***(iii)*** $V : \mathtt{P} \to \wp(W)$ is a *valuation function* (indicating the set of possible worlds in which each atom is true), and ***(iv)*** $A \subseteq \mathtt{P}$ is the *atomic awareness set* (indicating the set of atoms the agent is aware of). ◀

Note that *awareness-that* corresponds to the (local) neighbourhood function N and *awareness-of* is generated by the (global) atomic set A.

Definition 2 (Language) An *ANM* is described by the language $\mathcal{L}$, whose formulas φ, ψ are given by

$$\varphi, \psi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathrm{A}^{\mathrm{t}}\,\varphi \mid \mathrm{A}^{\mathrm{o}}\,\varphi \mid [*]\,\varphi,$$

with $p \in \mathtt{P}$. Formulas of the form $\mathrm{A}^{\mathrm{t}}\,\varphi$ are read as *"the agent is aware that φ"*, those of the form $\mathrm{A}^{\mathrm{o}}\,\varphi$ are read as *"the agent is aware of φ"*, and $[*]\,\varphi$ expresses that *"after the agent performs every possible deductive inference, φ holds"*. The set of atoms of any given $\varphi \in \mathcal{L}$, denoted by $\mathrm{atm}(\varphi)$, is defined in the standard way.[3] ◀

[3] That is, $\mathrm{atm}(\top) := \varnothing$, $\mathrm{atm}(p) := \{p\}$, $\mathrm{atm}(\neg\varphi) := \mathrm{atm}(\varphi)$, $\mathrm{atm}(\varphi \wedge \psi) := \mathrm{atm}(\varphi) \cup \mathrm{atm}(\psi)$, $\mathrm{atm}(\mathrm{A}^{\mathrm{t}}\,\varphi) := \mathrm{atm}(\varphi)$, $\mathrm{atm}(\mathrm{A}^{\mathrm{o}}\,\varphi) := \mathrm{atm}(\varphi)$, $\mathrm{atm}([*]\,\varphi) := \mathrm{atm}(\varphi)$.

Definition 3 (Semantic interpretation) The function $[\![\cdot]\!]^M : \mathcal{L} \to \wp(W)$, returning the set of worlds in a given *ANM* $M = \langle W, N, V, A \rangle$ in which a given φ is the case (the *truth-set of* φ in M), is defined inductively as follows. The cases for $\top$, atoms and Boolean operators are standard:

$$[\![\top]\!]^M := W, \quad [\![p]\!]^M := V(p), \quad [\![\neg\varphi]\!]^M := W \setminus [\![\varphi]\!]^M, \quad [\![\varphi \wedge \psi]\!]^M := [\![\varphi]\!]^M \cap [\![\psi]\!]^M.$$

For $\mathrm{A}^{\mathrm{o}}\,\varphi$, the formula is either globally true (when the agent is aware of all φ's atoms) or else globally false (otherwise):

$$[\![\mathrm{A}^{\mathrm{o}}\,\varphi]\!]^M := \begin{cases} W & \text{if } \mathrm{atm}(\varphi) \subseteq A; \\ \varnothing & \text{otherwise.} \end{cases}$$

For $\mathrm{A}^{\mathrm{t}}\,\varphi$, the formula is true at a world w in M if and only if the truth-set of φ in M is in the neighbourhood of w:

$$[\![\mathrm{A}^{\mathrm{t}}\,\varphi]\!]^M := \left\{ w \in W \mid [\![\varphi]\!]^M \in N(w) \right\}.$$

The remaining modality, $[*]$, is semantically interpreted not over M, but rather over its *augmentation*: the model M^* that results from making the neighbourhood of each world a set that contains the neighbourhood's *core* $\bigcap N(w)$ and is closed under supersets.[4] More precisely, from the given $M = \langle W, N, V, A \rangle$, define $M^* = \langle W, N^*, V, A \rangle$ with N^* such that

$$N^*(w) := \{ U \subseteq W \mid \bigcap N(w) \subseteq U \}.$$

Then,

$$[\![[*]\,\varphi]\!]^M := [\![\varphi]\!]^{M^*}.$$

As it will be recalled (subsection 3.2), in M^* the modality A^{t} behaves as $\Box$ does in relational models; this is the reason behind the intuitive reading of formulas of the form $[*]\,\varphi$, and the reason why $[*]$ is called here the *deductive closure* modality.

Satisfiability and validity are defined in the standard way, with the latter denoted as usual ($\Vdash \varphi$). ◀

3 Concepts, their properties and their relationship

The formal framework allows us to provide formal definitions for the notions the diagram on Figure 1 sketches.[5] Besides *awareness-of* and *awareness-that*, the concepts that will be discussed in detail are the following two, both falling under what the agent entertains:

- **Explicit knowledge:** $\mathrm{K}_{Ex}\,\varphi := \mathrm{A}^{\mathrm{o}}\,\varphi \wedge \mathrm{A}^{\mathrm{t}}\,\varphi$, what the agent is entertaining (is aware of) and has acknowledged as true (is aware that); this is what she 'really' knows.
- **Implicit knowledge:** $\mathrm{K}_{Im}\,\varphi := \mathrm{A}^{\mathrm{o}}\,\varphi \wedge [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$, what the agent is entertaining (is aware of) and will acknowledge as true after applying all possible deductive inferences.

[4] For specific details, see, e.g., Chellas (1980, Section 7.3).

[5] For a more detailed explanation of the theoretical framework, see Fernández-Fernández and Velázquez-Quesada (2019).

Note how one can also define further epistemic concepts, as the respective *awareness-of-less* counterparts of previous two. What the agent has acknowledged as true but is not currently entertaining (what will become explicitly known after she becomes aware of it), $\mathrm{K}_{Ex}^{-\mathrm{A}^{\mathrm{o}}}\,\varphi := \neg\,\mathrm{A}^{\mathrm{o}}\,\varphi \wedge \mathrm{A}^{\mathrm{t}}\,\varphi$, can be called *'disassociated' knowledge*; what she is not currently entertaining, and yet she can deduced from what she has acknowledged as true (i.e., what she could deduce after becoming aware of what she is aware that), $\mathrm{K}_{Im}^{-\mathrm{A}^{\mathrm{o}}}\,\varphi := \neg\,\mathrm{A}^{\mathrm{o}}\,\varphi \wedge [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$, can be called *currently 'unreachable' knowledge*.

3.1 Basic properties and relationships

It is now time to discuss the properties of the crucial concepts, and how they are related to their definitions elsewhere in the literature.

Awareness-of. The concept of *awareness-of*, A^{o}, is understood here as what the agent *entertains*. Thus, *awareness-of* is a matter of attention, and by itself it does not imply any attitude pro or con. Here, A^{o} is defined in terms of a global set of atomic propositions, the atomic awareness set A. As a consequence of this, *the agent is aware of the concept of 'truth'*: $\Vdash \mathrm{A}^{\mathrm{o}}\,\top$. This does not say that she is aware that this 'truth' holds everywhere; it simply states that she 'knows' that such concept exists. Technically, this is the case because $\top$, a primitive in the language[6], does not contain any atomic proposition.

Still, despite being aware of the concept of truth, the agent does not need to be aware of formulas that are always true: $\Vdash \varphi$ *does not imply* $\Vdash \mathrm{A}^{\mathrm{o}}\,\varphi$, and thus there might be validities the agent does not entertain. Technically, this is the case because a validity might have atoms (e.g., $p \vee \neg p$), and no atom is required to be in the atomic awareness set (e.g., p does not need to be in A).

On the other hand, *awareness-of* is defined not in terms of a set of formulas (as in the logic of general awareness by Fagin and Halpern) but rather in terms of a set of *atomic propositions*. As discussed already in Fagin and Halpern (1988), this makes the concept of *awareness-of closed under subformulas and superformulas*. More precisely,

$$\begin{array}{ll}
\Vdash \mathrm{A}^{\mathrm{o}}\,\neg\varphi \leftrightarrow \mathrm{A}^{\mathrm{o}}\,\varphi, & \Vdash \mathrm{A}^{\mathrm{o}}\,\mathrm{A}^{\mathrm{o}}\,\varphi \leftrightarrow \mathrm{A}^{\mathrm{o}}\,\varphi, \\
\Vdash \mathrm{A}^{\mathrm{o}}(\varphi \wedge \psi) \leftrightarrow (\mathrm{A}^{\mathrm{o}}\,\varphi \wedge \mathrm{A}^{\mathrm{o}}\,\psi), & \Vdash \mathrm{A}^{\mathrm{o}}\,\mathrm{A}^{\mathrm{t}}\,\varphi \leftrightarrow \mathrm{A}^{\mathrm{o}}\,\varphi, \\
 & \Vdash \mathrm{A}^{\mathrm{o}}\,[*]\,\varphi \leftrightarrow \mathrm{A}^{\mathrm{o}}\,\varphi.
\end{array}$$

In particular, note what the formulas on the right column indicate. The first states that awareness of awareness of a formula is equivalent to awareness of the formula. The second and the third indicate, with some paraphrasing, that entertaining φ is equivalent to entertaining the possibility of having accepted φ as true (the second) and also equivalent to entertaining the possibility of φ being true after deductive inference (the third).

An important difference between the *awareness-of* discussed here and the one in Fagin and Halpern (1988) is that, while the latter is local (what belongs to the awareness set assigned to the given evaluation point), the one proposed here is global (what belongs to the atomic awareness set assigned to *the whole model*). An intermediate position is held by Grossi and Velázquez-Quesada (2015), where (atom-based) *awareness-of* is defined as those formulas whose

[6] Not defined by an abbreviation of the form $p \vee \neg p$, as done in other proposals.

atoms appear in the awareness set of all worlds the agent cannot distinguish from the current evaluation point. Such definition makes sense in the multi-agent setting that the referred paper studies; in the single-agent case examined here, one can simply assume that the worlds in the model are exactly those that are relevant for the agent under discussion, thus making the two definitions conceptually equivalent.

Awareness-that. The concept of *awareness-that*, A^{t}, is understood here as what the agent has accepted/acknowledged as true; in this sense, it is a form of 'explicit' information. Still, it is not called *explicit knowledge* because the agent might not entertain such piece of information at the current stage. Thus, even though *awareness-that* is acknowledgement of truth, acceptance by itself does not imply that the agent is still entertaining such piece of information (she might have moved on to a different topic), and therefore it does not imply explicit knowledge either.

Here, A^{t} is defined as what appears in the neighbourhood of the current evaluation point.[7] In fact, the neighbourhood of a given world, a set of sets of worlds, can be understood as the list of formulas the agent has acknowledged as true, the important point being that these formulas are not represented syntactically (as a string of symbols), but rather semantically (as the set of worlds in the model in which the formula is true). Because of this purely semantic representation, the concept of *awareness-that* has an important closure property: $\Vdash \varphi \leftrightarrow \psi$ implies $\Vdash \mathrm{A}^{\mathrm{t}}\,\varphi \leftrightarrow \mathrm{A}^{\mathrm{t}}\,\psi$. In other words, the agent's *awareness-that* is *closed under logical equivalence*.[8] Thus, the agent is indeed omniscient in some sense as, at the level of acknowledgement, she cannot tell apart formulas that are true in exactly the same worlds in all models.[9]

Note that closure under logical equivalence is the only closure property the notion of *awareness-that* has. Different from other semantic representations of information, as the $\Box$-operator in standard epistemic logic under relational models, ***(i)*** $\Vdash \varphi$ does not imply $\Vdash \mathrm{A}^{\mathrm{t}}\,\varphi$, ***(ii)*** $\nVdash (\mathrm{A}^{\mathrm{t}}\,\varphi \wedge \mathrm{A}^{\mathrm{t}}\,\psi) \rightarrow \mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi)$, ***(iii)*** $\nVdash \mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow \mathrm{A}^{\mathrm{t}}\,\varphi$ and $\nVdash \mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow \mathrm{A}^{\mathrm{t}}\,\psi$. The reason for the failure of these properties is that no neighbourhood needs to have any closure property. In particular, ***(i)*** none of them needs to contain the whole domain (so $\mathcal{D}_M$, the truth-set of any validity, does not need to be in $N(w)$), ***(ii)*** none of them needs to be closed under intersections (so $[\![\varphi]\!]^M, [\![\psi]\!]^M \in N(w)$ does not imply $[\![\varphi]\!]^M \cap [\![\psi]\!]^M = [\![\varphi \wedge \psi]\!]^M \in N(w)$), and ***(iii)*** none of them needs to be closed under supersets (so $[\![\varphi \wedge \psi]\!]^M = [\![\varphi]\!]^M \cap [\![\psi]\!]^M \in N(w)$ implies neither $[\![\varphi]\!]^M \in N(w)$ nor $[\![\psi]\!]^M \in N(w)$). As a consequence of the latter two, *awareness-that* is not closed under modus ponens: $\nVdash \mathrm{A}^{\mathrm{t}}(\varphi \rightarrow \psi) \rightarrow (\mathrm{A}^{\mathrm{t}}\,\varphi \rightarrow \mathrm{A}^{\mathrm{t}}\,\psi)$.[10]

[7] In fact, A^{t}'s semantic interpretation is the 'set' semantic interpretation of the $\Box$-operator in neighbourhood models. (Recall: the 'subset' semantic interpretation makes $\Box\,\varphi$ true at w in M not only when $[\![\varphi]\!]^M$ is in $N(w)$, but also when any of its *subsets* is: $[\![\Box\,\varphi]\!]^M := \{w \in W \mid \text{there is } U \in N(w) \text{ such that } U \subseteq [\![\varphi]\!]^M\}$.)

[8] But, as it will be discussed later, this does not mean that the agent's *explicit knowledge* is closed under logical equivalence.

[9] This concept of *awareness that* (sometimes called *explicit knowledge*) lacks this property in other proposals, as Konolige (1984a) and Grossi and Velázquez-Quesada (2015), simply because they represent it as a set of formulas, and such set is not required to have any closure property.

[10] In this semantic setting, a modus ponens is a sequence of three steps: conjunction introduction to go from $[\![p]\!]^M$ and $[\![p \rightarrow q]\!]^M$ to $[\![p \wedge (p \rightarrow q)]\!]^M$, logical equivalence to go from the latter to $[\![p \wedge q]\!]^M$, and conjunction elimination to go from the latter to $[\![q]\!]^M$.

***Awareness-of* and *awareness-that*.** A property relating A^o and A^t has been discussed already ($\Vdash A^o A^t \varphi \leftrightarrow A^o \varphi$). Yet, for readers familiarised with awareness logic, the fact that *awareness-of* is global might suggest a further relationship between the two concepts: that the agent 'knows her own awareness'. Indeed, in the original Fagin and Halpern (1988), the fact that the awareness set of all worlds in the model is the same implies not only $\Vdash A\varphi \rightarrow \Box A\varphi$ (if the agent is aware of φ, then she knows this) but also $\Vdash \neg A\varphi \rightarrow \Box \neg A\varphi$ (if she is not aware of φ, then she knows this).

In the present setting, analogous properties do not need to hold: the agent does not need to acknowledge any formula, and thus she needs to acknowledge neither her awareness, $\nVdash A^o \varphi \rightarrow A^t A^o \varphi$, nor her unawareness, $\nVdash \neg A^o \varphi \rightarrow A^t \neg A^o \varphi$.

Explicit knowledge. The concept of *explicit knowledge*, K_{Ex}, is understood as those pieces of information the agent has acknowledged as true and is currently entertaining. Thus, explicit knowledge is what the agent is *aware-of* and *aware-that*: $K_{Ex}\varphi := A^o \varphi \wedge A^t \varphi$. Which are the consequences of this definition?

About *validities*. The agent does not need to know explicitly any validity, and none of the awareness concepts is, by itself, enough to guarantee that a validity is explicitly known: $\Vdash \varphi$ implies neither $\Vdash K_{Ex}\varphi$, nor $\Vdash A^o\varphi \rightarrow K_{Ex}\varphi$, nor $\Vdash A^t\varphi \rightarrow K_{Ex}\varphi$.

About *logical equivalence*. Even though *awareness-that* is closed under such property, A^o is not. Hence, *explicit knowledge* does not have this closure property: $\Vdash \varphi \leftrightarrow \psi$ does not imply $\Vdash K_{Ex}\varphi \leftrightarrow K_{Ex}\psi$. Still, *awareness-of* is the only piece that is missing. Thus, if two formulas are logically equivalent and the agent knows explicitly the first, awareness of the second gives the agent explicit knowledge of it: $\Vdash \varphi \leftrightarrow \psi$ implies $\Vdash (K_{Ex}\varphi \wedge A^o\psi) \rightarrow K_{Ex}\psi$.

About *closure under modus ponens*. *Awareness-that* lacks this property, and thus explicit knowledge lacks it too: $\nVdash K_{Ex}(\varphi \rightarrow \psi) \rightarrow (K_{Ex}\varphi \rightarrow K_{Ex}\psi)$. This is already shared by the explicit knowledge in Fagin and Halpern (1988) (which, recall, is defined as implicit knowledge, $\Box\varphi$, plus *awareness-of*, $A\varphi$). However, different from Fagin and Halpern's proposal, this proposal's additional requirement of being aware of the consequent does not give the agent explicit knowledge about it: $\nVdash K_{Ex}(\varphi \rightarrow \psi) \rightarrow ((K_{Ex}\varphi \wedge A^o\psi) \rightarrow K_{Ex}\psi)$. The agent might know explicitly an implication and its antecedent, and she might be entertaining the consequent, but still might fail to have explicit knowledge of the latter. What the agent is missing is realising that the consequent is actually the case:[11] $\Vdash K_{Ex}(\varphi \rightarrow \psi) \rightarrow ((K_{Ex}\varphi \wedge A^t\psi) \rightarrow K_{Ex}\psi)$.

3.2 Effects of the *augmentation* operation

The augmentation operation makes the neighbourhood of each world a set that contains the neighbourhood's core and is closed under supersets. Equivalently, one can understand the neighbourhood of each world after the augmentation operation as the result of adding the full domain to the original neighbourhood, and then close it under supersets and arbitrary intersections. It is well-known (e.g., Chellas 1980, Theorem 7.9) that in the resulting model, the *augmented*

[11] In this, the present setting coincides with Grossi and Velázquez-Quesada (2015).

model M^*, the operator A^{t} behaves as the standard $\Box$ does in relational models. Hence, occurrences of A^{t} under the scope of the modality $[*]$ can be understood as what the agent will acknowledge as true after she applies every possible deductive inference, making her *awareness-that* closed under logical consequence.[12] In this sense, $(\cdot)^*$ can be understood as a *full deductive inference* operation over A^{t}.

***Awareness-that* after deductive inference.** The operation $(\cdot)^*$ makes the agent's *awareness-that* closed under logical consequence; thus, after it, the agent acknowledges every validity: $\Vdash \varphi$ implies $\Vdash [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$. Moreover, after the operation, what the agent acknowledges is closed under both conjunction introduction, $\Vdash [*]((\mathrm{A}^{\mathrm{t}}\,\varphi \wedge \mathrm{A}^{\mathrm{t}}\,\psi) \rightarrow \mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi))$, and conjunction elimination, $\Vdash [*](\mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow \mathrm{A}^{\mathrm{t}}\,\varphi)$ and $\Vdash [*](\mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow \mathrm{A}^{\mathrm{t}}\,\psi)$. From the last two properties and the previous closure under logical equivalence, it follows that *awareness-that* is closed under modus ponens, $\Vdash [*](\mathrm{A}^{\mathrm{t}}(\varphi \rightarrow \psi) \rightarrow (\mathrm{A}^{\mathrm{t}}\,\varphi \rightarrow \mathrm{A}^{\mathrm{t}}\,\psi))$. The operation is a total function, so the last three properties can be described as $\Vdash ([*]\,\mathrm{A}^{\mathrm{t}}\,\varphi \wedge [*]\,\mathrm{A}^{\mathrm{t}}\,\psi) \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi)$, both $\Vdash [*]\,\mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$ and $\Vdash [*]\,\mathrm{A}^{\mathrm{t}}(\varphi \wedge \psi) \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\psi$, and $\Vdash [*]\,\mathrm{A}^{\mathrm{t}}(\varphi \rightarrow \psi) \rightarrow ([*]\,\mathrm{A}^{\mathrm{t}}\,\varphi \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\psi)$. Thus, the closure operation indeed makes A^{t} behave as $\Box$ under relational models.

Even more: after the operation the agent acknowledges her own *awareness-of*: $\Vdash [*](\mathrm{A}^{\mathrm{o}}\,\varphi \rightarrow \mathrm{A}^{\mathrm{t}}\,\mathrm{A}^{\mathrm{o}}\,\varphi)$ and $\Vdash [*](\neg\,\mathrm{A}^{\mathrm{o}}\,\varphi \rightarrow \mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{o}}\,\varphi)$ or, since the operation does not affect awareness sets, $\Vdash \mathrm{A}^{\mathrm{o}}\,\varphi \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\mathrm{A}^{\mathrm{o}}\,\varphi$ and $\Vdash \neg\,\mathrm{A}^{\mathrm{o}}\,\varphi \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{o}}\,\varphi$. Hence, the operation makes the relationship between A^{o} and A^{t} as it is in standard awareness models when the notion of *awareness-of* is global.

Implicit knowledge. This concept is defined here as what the agent currently entertains and will recognize as true *after* performing all possible deductive inferences, $\mathrm{K}_{Im}\,\varphi := \mathrm{A}^{\mathrm{o}}\,\varphi \wedge [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$.

For its properties, note how the agent does not need to know implicitly every validity, $\Vdash \varphi$ does not imply $\Vdash \mathrm{K}_{Im}\,\varphi$, the reason being that she might be unaware of (some of) the involved atoms.[13] Nevertheless, different from the *explicit knowledge* case, *awareness-of* is enough: the agent knows implicitly any validity she is currently entertaining: $\Vdash \varphi$ implies $\Vdash \mathrm{A}^{\mathrm{o}}\,\varphi \rightarrow \mathrm{K}_{Im}\,\varphi$.

The agent's implicit knowledge is *not closed under logical equivalence*, $\Vdash \varphi \leftrightarrow \psi$ does not imply $\Vdash \mathrm{K}_{Im}\,\varphi \leftrightarrow \mathrm{K}_{Im}\,\psi$, the *only* reason being, again, that the *awareness-of* requirement might fail (thus, $\Vdash \varphi \leftrightarrow \psi$ implies $\Vdash (\mathrm{K}_{Im}\,\varphi \wedge \mathrm{A}^{\mathrm{o}}\,\psi) \rightarrow \mathrm{K}_{Im}\,\psi)$. On the other hand, implicit knowledge is *closed under conjunction introduction* ($\Vdash (\mathrm{K}_{Im}\,\varphi \wedge \mathrm{K}_{Im}\,\psi) \rightarrow \mathrm{K}_{Im}(\varphi \wedge \psi)$) since, ***(i)*** after the operation, *awareness-that* has such property (here its 'distributed' version is the useful one), and ***(ii)*** *awareness-of* is closed under superformulas. It is also closed under *conjunction elimination* ($\Vdash \mathrm{K}_{Im}(\varphi \wedge \psi) \rightarrow \mathrm{K}_{Im}\,\varphi$ and $\Vdash \mathrm{K}_{Im}(\varphi \wedge \psi) \rightarrow \mathrm{K}_{Im}\,\psi$) as ***(i)*** after the operation, *awareness-that* has such property (again, its 'distributed'

[12] Velázquez-Quesada (2013) already explores this idea of using the known relationship between neighbourhood models and relational models to contrast knowledge before deductive reasoning with knowledge after. Still, the technical details are slightly different, as in the finite case that the referred paper studies, the augmented model is (equivalently) defined as the result of adding the full domain to each neighbourhood, and then closing it under *finite* intersections and supersets. Also, the referred paper's notion of explicit knowledge is different from the current proposal, as it does not take the concept of *awareness-of* into account.

[13] This is different from what happens with the same notion in Fagin and Halpern (1988), where the agent knows implicitly every validity.

version is the useful one), and ***(ii)*** *awareness-of* is closed under subformulas. These properties, together with the fact that *awareness-of* is closed under subformulas, tell us that implicit knowledge is *closed under modus ponens*: $\Vdash \mathrm{K}_{Im}(\varphi \rightarrow \psi) \rightarrow (\mathrm{K}_{Im}\,\varphi \rightarrow \mathrm{K}_{Im}\,\psi)$.

Notice how, while the implicit knowledge of Fagin and Halpern (1988) contains all validities and is closed under logical consequence, implicit knowledge as defined here might not contain all validities and, despite being closed under modus ponens, it does not need to be closed under logical equivalence. This is because, while implicit knowledge in Fagin and Halpern (1988) is the agent's 'semantic' information (given by the modal operator $\Box$), here it is the closure under modus ponens of what the agent has acknowledged as true ($[*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$) and is currently entertaining ($\mathrm{A}^{\mathrm{o}}\,\varphi$); it is the closure under modus ponens of what the agent knows explicitly. This highlights the crucial difference between the understanding of explicit knowledge in both settings. In Fagin and Halpern (1988), what is needed for implicit knowledge to be explicit is for the agent to be aware of the given formula. However, in this proposal, knowing a formula implicitly already makes the agent aware of it. Thus, here, what is needed for implicit knowledge to become explicit is not an act of awareness raising; what is needed is rather an act of *deductive inference*.

Moorean phenomena. In proposals dealing with implicit and explicit knowledge, a particular property is recurrent: explicit knowledge is also implicit knowledge. In Fagin and Halpern (1988), this follows from the fact that explicit knowledge is defined as the implicit knowledge that satisfies an additional requirement (*awareness-of*); in settings distinguishing explicit information from implicit one by means of deductive reasoning (e.g., Konolige 1986), this follows from the fact that deductive inference is monotone.

This property, seemingly not only natural but rather essential, is not satisfied by the presented setting: $\nVdash \mathrm{K}_{Ex}\,\varphi \rightarrow \mathrm{K}_{Im}\,\varphi$. The reason is, as shown in Velázquez-Quesada (2013), that what the agent has acknowledged as true at some particular stage does not need to be acknowledged as true after the augmentation operation, that is,

Fact 1 $\nVdash \mathrm{A}^{\mathrm{t}}\,\varphi \rightarrow [*]\,\mathrm{A}^{\mathrm{t}}\,\varphi$

Proof. Take $\varphi := \neg\,\mathrm{A}^{\mathrm{t}}\,q$, and consider $M = \langle W = \{w_1, w_2, w_3, w_4\}, N, V, \varnothing\rangle$, a model with four worlds over the set of atomic propositions $\{p, q\}$, with $V(p) = \{w_1, w_2\}$ and $V(q) = \{w_1, w_3\}$. The *awareness-of* set A is not relevant (hence empty, for simplicity), but the neighbourhood function is: suppose it is given by $N(w_1) := \{\{w_1, w_2\}, \{w_1, w_3, w_4\}, W\}$ and $N(w_2) = N(w_3) = N(w_4) := \varnothing$. Since $[\![\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M} = W$ is in $N(w_1)$, we have $w_1 \in [\![\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M}$.

However, the neighbourhood function of the augmented model, N^*, is such that $N^*(w_1) = \{U \subseteq W \mid \{w_1\} \subseteq U\}$ and $N^*(w_2) = N^*(w_3) = N^*(w_4) = \{W\}$. Note how $[\![\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M^*} = \{w_2, w_3, w_4\}$ is *not* in $N^*(w_1)$, so $w_1 \notin [\![\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M^*}$, that is, $w_1 \notin [\![[*]\,\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M}$. ∎

Thus, while in M the agent has acknowledged $\neg\,\mathrm{A}^{\mathrm{t}}\,q$ as true at w_1 (i.e., $w_1 \in [\![\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M}$), the operation changes this: in M^*, the agent has not acknowledged $\neg\,\mathrm{A}^{\mathrm{t}}\,q$ as true at w_1 (i.e., $w_1 \notin [\![\mathrm{A}^{\mathrm{t}}\,\neg\,\mathrm{A}^{\mathrm{t}}\,q]\!]^{M^*}$). One just needs to make the agent aware of the involved formula $\neg\,\mathrm{A}^{\mathrm{t}}\,q$ (e.g., take $A := \{p, q\}$) to obtain a model (M) and a world (w_1) in which the agent knows a formula ($\neg\,\mathrm{A}^{\mathrm{t}}\,q$) explicitly,

$$w_1 \in [\![\mathrm{K}_{Ex}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M = [\![\mathrm{A}^{\mathrm{o}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q \;\wedge\; \mathrm{A}^{\mathrm{t}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M$$
$$= [\![\mathrm{A}^{\mathrm{o}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M \cap [\![\mathrm{A}^{\mathrm{t}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M,$$

and yet she does not know it implicitly,

$$w_1 \notin [\![\mathrm{K}_{Im}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M = [\![\mathrm{A}^{\mathrm{o}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q \;\wedge\; [*]\, \mathrm{A}^{\mathrm{t}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M$$
$$= [\![\mathrm{A}^{\mathrm{o}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M \cap [\![[*]\, \mathrm{A}^{\mathrm{t}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q]\!]^M.$$

So, is there some fundamental problem with the current proposal? To answer this, first note how the 'explicit is implicit' property does not always fail. In fact, it holds for a large class of formulas, including not only the purely propositional ones, but also all those whose truth-set *does not shrink* as a consequence of the augmentation operation:

Proposition 3.1 *$\Vdash \varphi \rightarrow [*]\,\varphi$ implies $\Vdash \mathrm{K}_{Ex}\,\varphi \rightarrow \mathrm{K}_{Im}\,\varphi$.*

Proof. See that of Velázquez-Quesada (2013, Proposition 2). ■

Then, why do the rest of the formulas fail? The provided counterexample, $\neg\, \mathrm{A}^{\mathrm{t}}\, q$, shows one of their crucial feature: they express not ontic facts, but rather *epistemic* situations and, in particular *negative awareness-that* situations. Indeed, $\neg\, \mathrm{A}^{\mathrm{t}}\, q$ expresses that the agent has not acknowledged q as true, and then $\mathrm{A}^{\mathrm{t}}\, \neg\, \mathrm{A}^{\mathrm{t}}\, q$ says that the agent has acknowledged this. In other words, and considering her *awareness-of*, the agent knows explicitly that she does not know q explicitly.

However, the agent might have enough information to get to know what she currently knows she does not have. Indeed, in the provided model (and with $A := \{p, q\}$), she knows explicitly both p and $p \rightarrow q$ (both $[\![p]\!]^M = \{w_1, w_2\}$ and $[\![p \rightarrow q]\!]^M = \{w_1, w_3, w_4\}$ are in $N(w_1)$). Then, after deductive reasoning, she will realise that q is indeed the case, hence knowing q explicitly; but then, she will automatically stop acknowledging (and thus stop knowing explicitly) that she did not know q explicitly. In other words, she might know explicitly that she does not know q, but such high-order knowledge will be gone once she gets to know that q is indeed the case.

The reason for the failure of the 'explicit is implicit' property is that the agent has knowledge not only about propositional facts but also about her own (and eventually other agents') knowledge. This knowledge (semantically, the neighbourhood function) changes through the closure operation, so the agent might know something explicitly at some point, and yet not know it explicitly after (semantically, the *awareness-that* component is the key: we might have $U \in N(w)$ with $U = [\![\varphi]\!]^M$ for some φ, but even though this implies $U \in N^*(w)$, nothing guarantees $U = [\![\varphi]\!]^{M^*}$). This is an instance of the so called 'Moorean phenomena', which occurs when an epistemic action invalidates itself. In its best known incarnation, this phenomenon appears as formulas that become false after being truthfully announced (van Ditmarsch and Kooi 2006, Holliday and Icard 2010); here, it appears as formulas that become false after deductive inference.

4 Summary and ongoing work

We have seen so far our model for representing the concept of *explicit knowledge*, based on the combination of *awareness that* and *awareness of*. Afterwards we have shown some properties of these concepts that highlight not only the relationship between them, but also some of the advantages our model may have with respect to alternative proposals such as the logic of awareness (Fagin and Halpern 1988) or deduction systems (Konolige 1984a).

What our system is still missing is a further exploration into the dynamics, i.e., we need to define further epistemic actions that will show how the information changes throughout the process of gaining or loosing knowledge. Actions like becoming (un)aware, performing a single-step deductive inference or observing a piece of information will be crucial for this. We would also like to complete a comprehensive comparison for the notions of *awareness of* and *awareness that* with similar proposals, such as the concept of *topics* in Berto and Hawke (2018) for the former, or the proposals of *explicit knowledge* that do not incorporate the awareness of the agent for the latter, like Konolige (1984a), Artëmov and Nogina (2005) or Velázquez-Quesada (2013).

For future work we leave the axiomatization of the framework and its epistemic actions, together with the modelling of the intuitive interpretation of the concept of justification, understood here as the actions that turn information into explicit knowledge.

References

S. N. Artëmov and E. Nogina. Introducing justification into epistemic logic. *Journal of Logic and Computation*, 15(6):1059–1073, 2005. doi: 10.1093/logcom/exi053.

F. Berto and P. Hawke. Knowability relative to information. *Mind*, Oct. 2018. doi: 10.1093/mind/fzy045. url: https://philpapers.org/go.pl?id=6248898&aid=BERKRTv1.

B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, Mass., 1980. ISBN 0-521-295157. doi: 10.1017/CBO9780511621192.

H. van Ditmarsch and B. P. Kooi. The secret of my success. *Synthese*, 151(2): 201–232, 2006. doi: 10.1007/s11229-005-3384-9.

F. Dretske. Conscious experience. *Mind*, 102(406):263–283, Apr. 1993. doi: 10.1093/mind/102.406.263.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988. ISSN 0004-3702. doi: 10.1016/0004-3702(87)90003-8.

C. Fernández-Fernández and F. R. Velázquez-Quesada. A formal model for explicit knowledge as awareness-of plus awareness-that. In I. Sedár and M. Blicha, editors, *The Logica Yearbook 2018*, pages 101–115. College Publications, London, UK, 2019. ISBN 978-1-84890-307-4. url: http://bit.ly/2mm6sq1.

D. Grossi and F. R. Velázquez-Quesada. Syntactic awareness in logical dynamics. *Synthese*, 192(12):4071–4105, Dec. 2015. ISSN 0039-7857. doi: 10.1007/s11229-015-0733-1. url: http://bit.ly/1Ff7a55.

J. Y. Halpern and L. C. Rêgo. Interactive unawareness revisited. *Games and Economic Behavior*, 62(1):232–262, 2008. ISSN 0899-8256. DOI: 10.1016/j.geb.2007.01.012.

A. Heifetz, M. Meier, and B. C. Schipper. A canonical model for interactive unawareness. *Games and Economic Behavior*, 62(1):304–324, 2008. DOI: 10.1016/j.geb.2007.07.003.

J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, N.Y., 1962. ISBN 1-904987-08-7.

W. H. Holliday and T. F. Icard. Moorean phenomena in epistemic logic. In L. Beklemishev, V. Goranko, and V. Shehtman, editors, *Advances in Modal Logic*, pages 178–199. College Publications, 2010. ISBN 978-1-84890-013-4. URL: `http://www.aiml.net/volumes/volume8/Holliday-Icard.pdf`.

K. Konolige. Belief and incompleteness. Technical Report 319, SRI International, 1984a.

K. Konolige. *A Deduction Model of Belief and its Logics*. PhD thesis, Computer Science Department, Stanford University, Stanford, USA, 1984b.

K. Konolige. What awareness isn't: A sentential view of implicit and explicit belief. In J. Y. Halpern, editor, *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge, Monterey, CA, March 1986*, pages 241–250. Morgan Kaufmann, 1986. URL: `http://www.tark.org/proceedings/tark_mar19_86/p241-konolige.pdf`.

R. Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970. DOI: 10.1111/j.1755-2567.1970.tb00434.x.

E. Pacuit. *Neighborhood Semantics for Modal Logic*. Springer, 2017. ISBN 978-3-319-67148-2. DOI: 10.1007/978-3-319-67149-9. URL: `http://web.pacuit.org/files/neighborhoods/nbhd-v6-PENULTIMATE.pdf`.

B. C. Schipper. Awareness. In H. van Ditmarsch, J. Y. Halpern, W. van der Hoek, and B. Kooi, editors, *Handbook of Epistemic Logic*. College Publications, London, 2015. ISBN 978-1-84890-158-2. DOI: 10.2139/ssrn.2401352.

D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, pages 143–173. Reidel, Dordrecht, The Netherlands, 1970. ISBN 978-94-010-3272-8. DOI: 10.1007/978-94-010-3272-8˙7.

R. Stalnaker. The problem of logical omniscience, I. *Synthese*, 89(3):425–440, 1991. ISSN 0039-7857. DOI: 10.1007/BF00413506.

F. R. Velázquez-Quesada. Explicit and implicit knowledge in neighbourhood models. In D. Grossi, O. Roy, and H. Huang, editors, *LORI*, volume 8196 of *Lecture Notes in Computer Science*, pages 239–252. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-40947-9. DOI: 10.1007/978-3-642-40948-6˙19. URL: `http://bit.ly/1yIRFF5`.