What do epistemic logic and cognitive science have to do with each other?

Ahti-Veikko Pietarinen
Department of Philosophy, University of Helsinki, P.O. Box 9, FIN-00014 Helsinki, Finland

Abstract

Epistemic logic is a multi-faceted theory aimed at targeting notions such as knowledge, belief, information, awareness, memory and other propositional attitudes, by means of logical and semantical tools. These concepts ought to be in the spotlight of cognitive science too, but the two have not yet seriously been explored in cooperation. In this paper, it is shown that a number of possibilities is opened up by attempting to answer the question of what epistemic logic and cognitive science have to do with each other. Among the proposed answers are: (i) new quantified versions of multi-agent epistemic logic capture locutions involving object identification, giving rise to applications in representing knowledge in multi-agent systems and parallel processing. (ii) The framework of game-theoretic semantics for the ensuing logics enjoys increased cognitive plausibility as the true semantics for epistemic notions. (iii) Several recent findings in cognitive neuroscience pertaining to the notions of awareness and explicit versus implicit processing contribute to logical studies. These three connections are explored here from both logical and cognitive perspectives. Reflecting neuroscientific research, new extensions of epistemic logic are defined, increasing formal understanding of unconscious and unaware information processing in the brain, and making the formalism thus amenable to knowledge representation in multi-agent configurations.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Epistemic logic; Knowledge; Multi-agent systems; Game-theoretic semantics; Cognitive neuroscience; Awareness

1. Introduction: epistemic logic and cognitive science

1.1. Motivation and rationale of the paper

Recent years have witnessed a growing interest in many varieties of logics of propositional attitudes, including those employing notions of knowledge and belief. These are collectively known as epistemic logics. The growth has largely been due to increased interest in epistemic aspects of multi-agent systems and computational questions concerning the underlying logics, typically needed in computer science and artificial intelligence (Fagin, Halpern, Moses, & Vardi, 1995). To compensate this dominance of computational concerns, it is time to call for a renewed exploration of epistemic logic from the perspectives of cognitive science and the related neuroscientific fields too.

However, in cognitive sciences, the motivations...
and desiderata have to be seen as somewhat different from those of computing or artificial intelligence. Notions such as expressive power and the variability in the meaning of knowledge warrant a more prominent place, following the venerable philosophical tradition. However, upon a closer inspection, the interplay of epistemic logic and cognitive science is seen to yield to a couple of new, till now largely unnoticed opportunities. It turns out that the findings in cognitive science complement many of those questions asked in logic. To do this, we call attention to the modelling potentialities of epistemic languages in the light of recent findings in cognitive neuroscience. These findings are so pressing as to warrant one to develop renewed cognitively justified approaches not only to logics of propositional attitudes but also to the overall field of logical semantics.

Cognitive science has not yet met these challenges, however. The extant epistemic systems borrowed from computer science are still of relatively poor expressive power, and usually just propositional languages are being employed. Thus they are unlikely candidates to be genuinely interesting beyond computational concerns.

What are the desiderata for epistemic logic that will display increased cognitive value? The following items delineate some of the issues that need to be kept in mind in the interplay between epistemic logic and cognitive science. (i) To ensure enough expressive power, quantified notions should be incorporated into epistemic logic. (ii) As useful applications have dealt increasingly with multi-agent systems, and as epistemic logic (as originally conceived) lends itself well to such multi-agent considerations, such applications should be pursued with these new developments in view. (iii) The semantic framework for epistemic notions should itself present genuine cognitive and neuroscientific plausibility.

With respect to these three items, in this paper I seek to establish three motivational backings. (1) By means of quantifiers, one can ask interesting questions concerning the identity of objects. Various ways of identifying objects are of direct interest in cognitive science, and they interlock with the game-theoretic approach to logical semantics that will be used in this paper. Contrary to the claims persisting in the tradition of quantified epistemic logic (and quantification in modal logic in general), such extensions are not particularly problematic. (2) By going from perfect, classical languages to those involving imperfect information one can hope to develop more realistic logics of knowledge and other attitudes. These notions can be made more precise by relating the semantic notions of information to games in the sense of the theory of games (von Neumann & Morgenstern, 1944). (3) As cognitive neuroscience has provided interesting new insights into notions such as implicit versus explicit information processing in tasks like perception and identification, memory and its retrieval, and belief content and belief formation, it has brought the notion of awareness at the forefront of human knowledge. The logics suggested here reflect these insights by incorporating new operators into the language and making good use of the associated semantic frameworks, namely the game-theoretic and possible-worlds ones. The amalgamation of these can in fact be seen to implement interesting cognitive notions such as strategic decisions, memory and recall, and forms of bounded rationality.

Furthermore, the items (i)–(iii) are related in the following sense: the semantic tool that is used, namely the game-theoretic semantics of imperfect information, shows how specific identification (knowledge de re) can be attained. In order to have specificity, as in locutions involving direct knowledge plus wh-constructions (knowing who, what, where, when, whether etc.), one must be able to maintain that knowledge across varying situations, changing locations and times. The general concept here is that of a possible world. This observation can be made rigorous by observing the structural similarities between extensive forms of a semantic game and the possible-worlds semantics for knowledge.

Apart from this result, there are several conceptual and interpretational issues to be addressed. Epistemic logic itself can be harnessed to model cognitive phenomena, including the distinction of explicit versus implicit knowledge and memory, different senses of awareness, and other propositional attitudes such as belief and perception. In this paper, we draw logical distinctions that reflect, or can be made to reflect, many recent neuroscientific findings. That is, new ways of extending the basic epistemic language to cover empirically supported phenomena in cognitive neuroscience are presented, including blindsight, unilateral neglect, and aspects of amnesia and memory. Related cognitive phenomena are in fact over-
whelming, and several guidelines for future work will be drawn up here.

In addition to cognitive science, knowledge representation and multi-agent systems benefit from these renewed formalisms. To this effect, a case study is presented that aims at showing that these more expressive formalisms are in fact indispensable in properly representing knowledge in multi-agent systems and concurrent processors communicating with one another.

Hintikka (1989) was perhaps the first publication to recognise aspects of the interrelations between experimental findings in neuroscience and logic, including important philosophical perspectives to Cartesian notions of the famous cogito argument. One of the primary insights of Hintikka’s paper was to differentiate between two modes of identification, perspective versus public one, and to relate that distinction to neuroscientific and cognitive correlate of ‘where’ versus ‘what’ systems, the distinction that has been observed to be relevant to higher-level cognitive functions (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982; Vaina, 1987). This paper is complementary to some of Hintikka’s remarks. The aim is to go on to formalise ideas that neuroscientific research has provided us with in recent years by semantical and logical tools derived from the work that has been done on epistemic aspects of propositional attitudes. The distinction between different modes of identification remains valid, but for the purposes of this paper the kinds of propositional epistemic logic that are our primary tools here are not in general strong enough to bring out this distinction. The distinction between perspectival versus public systems of identification raises its head in the quantified cases that will be exposed only briefly. The point (i) above is in essence assumed to subsume this distinction.

The distinction is also relevant to what is termed focused knowledge in this paper, and to the overall implicit/explicit distinction showing up both in neuroscientific experiments and in epistemic logic.

1.2. Brief introduction to classical epistemic logic

According to the Greeks, knowledge meant true, justified belief. Hintikka (1962) proposed that knowledge could be viewed as truth in epistemic alternatives. Many names can be given to these epistemic alternatives: they can be possible worlds, alternative states, scenarios, points in time or space, and so on. We prefer possible worlds, but shall use the term state and the term world interchangeably. The key insight is that an agent i knows a proposition p in the state or world w if and only if p is true in all of i’s epistemic alternatives to w.

This analysis has led to a proliferation of knowledge in logic, philosophy, computer science, artificial intelligence and economics (interactive epistemology). Nowadays one finds notions such as common, shared and distributed knowledge and belief. Furthermore, the distinction between knowledge de dicto and knowledge de re in quantified versions of epistemic logic is widely spread in the philosophy of language. The revival of Hintikka’s paradigm is largely due to the series of discoveries and discoveries made for the purposes of computer science, many of them collected in Fagin et al. (1995).

The well-formed formulas of ordinary propositional epistemic logic L can be defined as follows:

\[
\phi::= p | \varphi \lor \psi | \neg \varphi | K_i \varphi.
\]

\(K_i \varphi\) is read ‘an agent i knows \(\varphi\)’, where \(\varphi\) is an arbitrary formula of the language. Alternatively, in some cases one can speak of merely information, and translate this as ‘i is informed about \(\varphi\)’.

The semantics is based on possible worlds along the following lines. Let \(\varphi, \psi\) be formulas of ordinary propositional epistemic logic L. A model is a triple \(M = (W, R, g)\), where g is a total valuation function \(g: W \rightarrow (\Phi \rightarrow \{\text{True}, \text{False}\})\), assigning to each proposition letter \(p \in \Phi\) a subset of a set of possible worlds \(W' = \{w_0, w_1, \ldots\}\) such that \(|w_i| g(p) = \text{True}, w \in W\). The component \(R = \{r_1, \ldots, r_n\}\) denotes a set of accessibility relations for each agent \(i = 1 \ldots n, r_i \subseteq W' \times W\). Let us write \(w_i \in [w_0]_{r_i}\) to denote that a possible world \(w_i\) is \(i\)-accessible from \(w_0\), that is, \(w_i\) belongs to the set of \(i\)'s epistemic alternatives.

In order to see when a formula of this language is true in a model \(M\) and a world \(w\), the truth-clauses are defined in the following customary way:

\[
(M, w) \models p \iff g(w)(p) = \text{True}
\]

\[
(M, w) \models \neg \varphi \iff (M, w) \not\models \varphi
\]

\[
(M, w) \models \varphi \lor \psi \iff (M, w) \models \varphi \text{ or } (M, w) \models \psi
\]

\[
(M, w) \models K_i \varphi \iff (M, w') \models \varphi, \text{ for all } w' \in [w]_{r_i}.
\]
Hence, $K_i \varphi$ is true in $(M, w)$ whenever $\varphi$ is true in all possible worlds that are $i$-accessible from $w$.

The basic strategy is thus to analyse knowledge as an attitude of the subject towards a proposition. The sentence ‘I know that a cat is on the mat’ consists of one agent, $i$, meaning here the first-person reflexive, a proposition $\varphi = 'a$ cat is on the mat’, and a relation between the agent $i$ and a proposition. Semantically, this means that to know something is to realise what alternative states or possible worlds it includes and also what it excludes. The more the agent knows or is being informed about the less alternative states there are where the objects of knowledge, or propositions, will hold. For the example sentence ‘I know that a cat is on the mat’ to represent a true statement, the expression, of which the left-hand side operator $\varphi$ holds. For the example sentence ‘I know that a cat is on the mat’ consists of an $i$- accessible from $w_0$ where the agent $i$ subsists, while those alternative states where it is not the case that the cat is on the mat are not linked by the relation $\rho_i$ with the actual world.

Given these starting points, let us now turn to a couple of new ways of extending this basic insight to knowledge.

### 2. Extension I: from perfect to imperfect information

The language above is, like typically most received classical logics, a logic of perfect information transmission. Syntactically, this means that the formulas are written using a linear form of symbolisation. Semantically, the information transmission is in a certain well-defined sense an uninterrupted one.

These assumptions are inadequate for many purposes. As an example, in Section 6 it is demonstrated that logical modelling of knowledge needed in understanding certain communicating multi-agent systems requires much more expressive formalisms where these assumptions are dropped. In order to capture logical and semantic ideas of imperfect information, somewhat refined epistemic languages are needed.

Hence, the first extension is this. Let $K_i \psi$ be an $\mathcal{L}$-formula, and let $A=\{K_1, \ldots, K_n\}$, $K_i \in A$, $i \in \{1 \ldots n\}$ such that $K_j$ is in the scope of $K_1 \ldots K_n$. Then:

- if $B \subseteq A$, then $(K_i/B) \psi \in \mathcal{L}^*$, $K_i \in B$.

For example, $K_1(K_2/K_1) \varphi$ and $K_1(\varphi \land (K_2/K_1) \psi)$ are well-formed formulas of $\mathcal{L}^*$. As an example, $K_1(K_2/K_1) \varphi$ can be read as ‘1 and 2 know that $\varphi$’. It is both linguistically and logically taken to be different from ‘1 knows that $\varphi$ and 2 knows that $\varphi$’. That is, $K_1(K_2/K_1) \varphi$ does not reduce to the conjunctive proposition $K_1 \varphi \land K_2 \varphi$. The purpose of the backslash notation is to denote where perfect information as required per classical notation breaks down and will be turned into imperfect information.

We read it so that the operators on the right-hand side of the slash are those occurring in the context of the expression, of which the left-hand side operator is to be semantically independent. The meaning of these formulas will be given by semantic games in Section 3.

The second step is to add quantifiers to the propositional base language. This can be done either by adding quantifiers to perfect information language $\mathcal{L}$ or to the already extended $\mathcal{L}^*$. It is advisable to take the first-order epistemic logic $\mathcal{L}_{wau}$ as the starting point and extend it by applying the slash notation to it. Thus, let $\mathcal{L}_{wau}$ consist of a signature $\tau$, a logical vocabulary, and the following way to build up formulas:

$$ \phi ::= S[K_i \varphi] \forall x \exists y \varphi | \psi \lor \varphi | \neg \varphi | x = y. $$

Now we will also have universal and existential quantifiers, plus a two-place identity sign.

Then, let $Q \psi$, $Q \in \{\forall x, \exists y, K_i\}$ be an $\mathcal{L}_{wau}$-formula in the syntactic scope of the elements in $A = \{K_1, \ldots, K_n, \forall x_i, \exists y_i\}$ (for finite integers $i$, $j$, $k$, $n$; $j \neq k$). Then $\mathcal{L}_{wau}^*$ consists of the wffs of $\mathcal{L}_{wau}$ together with the rule:

- if $B \subseteq A$, then $(Q/B) \psi$ is an $\mathcal{L}_{wau}^*$-formula, $Q \in B$.

For example, in the $\mathcal{L}_{wau}^*$-formula

$$ K_{v_j} \exists y(\exists x/K_{v_i}, y)(K_{v_j}/K_{v_i}, y)S_{xy}, $$

the information about the choices for $K_{v_i}$ and $y$ is hidden in these positions where $\exists x$ and $K_{v_i}$ are evaluated. I will return to the meaning of this kind of
quantification in later sections. In the next section, we define the semantics for $\mathcal{L}^*$ by means of games.

3. Extension II: from static semantics to dynamic games

Unfortunately, the classical possible-worlds semantics as outlined in Section 1.2 is static in the sense that it cannot account for the dynamics of what is going on in the evaluation of the formulas of these new languages. One particular impediment is that $\mathcal{L}^*$ or, for that matter, $\mathcal{L}^*_{\text{non}}$ cannot assume just the traditional possible-worlds semantics, because there may be some subformulas that do not receive meaning when severed from their context. For instance, the formula $(K_{V_1} / K_{V_2}) \psi$ does not have any interpretation in classical possible-worlds semantics. For these reasons, we define semantics non-compositionally with the aid of games.

For basic ideas of game-theoretic semantics, see e.g. Hintikka and Sandu (1997), Pietarinen (2003b).\footnote{Pietarinen (2003c) is a more generic survey of game theories across formal scientific inquiry.}

The game-theoretic semantics for $\mathcal{L}^*$ is as follows. Every $\mathcal{L}^*$-formula $\varphi$ defines a game $\mathcal{G}(\varphi, w, M)$ on a model $M$ between two finite teams of players, the team of falsifiers $F = \{F_1, \ldots, F_l\}$ and the team of verifiers $V = \{V_1, \ldots, V_k\}$. $w$ is a possible world, and $g$ is an assignment to the propositional letters as in the classical case. The semantic game $\mathcal{G}(\varphi, w, M)$ is then defined by the following rules.

(G.atom): if $S$ is atomic, the game terminates. The team $V$ wins the play of the game if $S$ is true in a given interpretation, and the team $F$ wins the play of the game if $S$ is false in a given interpretation.

(G,$\neg$): if $\varphi = \neg \psi$, then the players $V$ and $F$ change their roles, that is, $V$ becomes $F$ and $F$ becomes $V$. The winning conventions of (G.atom) will also change throughout the rest of the game. The next choice is made in $\mathcal{G}(\psi, w, M)$.

(G,$\lor$): if $\varphi = (\psi \lor \theta)$, then $V$ chooses from the set $\{\text{Left}, \text{Right}\}$, and the next choice is made in $\mathcal{G}(\psi, w, M)$ if the choice was Left, and in $\mathcal{G}(\theta, w, M)$ if the choice was Right.

(G,$K_i$): if $\varphi = K_i \psi$ and the game has reached $w$, then a member $F_j \in F$ chooses $w_1 \in [w]_{\psi}$, and the next choice is made in $\mathcal{G}(\psi, w_1, M)$.

(G,$K_i / B$): if $\varphi = (K_i / B) \psi$, $K_i \notin B$ and the game has reached $w$, then a member $F_j \in F$ chooses $w_1 \in \mathcal{W}$ independently of the choices made for the elements in $B$, and the next choice is made in $\mathcal{G}(\psi, w_1, M)$.

The formulas $(K_i / B) \psi$ mean that there is imperfect information in the sense that the player or the team member $F_j$ choosing for $K_i$ on the left-hand side of the slash is not informed about the choices made for the elements in $B$ earlier in the game. The terminology of using teams and members of the team is now explained by the fact that in the game rule $(G,K_i / B)$, as soon as the same player, namely the falsifier, has been involved in the choices related to the elements in $B$, he is seen to forget information. This is implemented by taking there to be individual members of the falsifier team whose information is persistent but who do not share this information with $F$. In this way $F$, who can also be seen as the principal coordinator deciding whether new members are needed to act, may forget information about previous choices.

Let us note two things. Since we want to leave the final interpretation of these modalities somewhat open, no mention is made in the last rule of the accessibility relation. Secondly, knowledge of players that crops up in the semantic game $\mathcal{G}(\varphi, w, M)$ should not be confused with that of knowledge of agents of the object language.

The purpose of the team $V$ is to show that $\varphi$ is true in a possible-worlds model $M$ at a world $w$. Dually, the purpose of the team $F$ is to show that $\varphi$ is false in a model $M$ at $w$. The winning conventions are defined so that if the atomic formula $p$ is true according to the given valuation $g$ then $V$ wins, and if $p$ is false according to the given valuation $g$ then $F$ wins.

To attain the truth and falsity of formulas $\varphi$, the concept of a strategy is used. Given a game $\mathcal{G}(\varphi, w, M)$, a strategy for the teams $V$ or $F$ is a function assigning to each non-atomic subformula a member of a team, outputting an application of each rule, which is a possible world, a value in $\{\text{Left}, \text{Right}\}$ (as prompted by logical connectives), or an instruc-
tion to change roles (as prompted by negation). A winning strategy is now a set of strategies by which teams can make operational choices such that every play results in a win for him or her, no matter how the opponent chooses.

The truth is now defined as follows. Let \( \varphi \) be an \( \mathcal{L}^* \)-formula. For any model \( M \), a valuation \( g \), and \( w \in \mathcal{W} \), \( \varphi \) is true in \( M \) if and only if a strategy exists that is winning for the player who initiated the game \( \mathcal{R}(\varphi, w, M) \) as the player \( V \). Dually, \( \varphi \) is false in \( M \) if and only if a strategy exists that is winning for the player who initiated the game \( \mathcal{R}(\varphi, w, M) \) as the player \( F \).

The distinguishing feature of semantic games is that the direction of the flow of information is reversed from that of classical semantics: one starts the evaluation with an outermost component in the whole formula and proceeds outside in, towards contextually constrained subexpressions. These games give rise to a diagrammatic system in the sense that is explained in Section 7, where imperfect information can be captured by equivalence relations.

A game \( \mathcal{G} \) is determined just in case for every play on \( \varphi \), either \( V \) has a winning strategy in \( \mathcal{G} \) or \( F \) has a winning strategy in it. It is easy to see that games for \( \mathcal{L}^* \) are not determined: there are formulas whose correlated games cannot be won by either player, no matter what the strategies they use.

From non-determinacy it follows that the law of excluded middle (\( \varphi \lor \neg \varphi \)) fails in these logics. This is a common thing to happen whenever there is imperfect information. Further, non-determinacy is closely related to that of partiality (Section 14).

4. Quantified epistemic logic not always linear

Technically, our quantified epistemic language \( \mathcal{L}_{\mathfrak{w}}^* \) has a novelty of building up two-dimensional formulas with partially ordered quantifier-operator structures. A simple such formula can be written as:

\[
\exists x \quad K_{v_1} \quad \exists y \quad Sxy.
\]

We will avail ourselves of the possibilities opened up by these novelties as we proceed. The meaning of these formulas comes from extended game-theoretic semantics (Section 7).

By resorting to the slash notation we can linearise (2) to the logically equivalent \( \mathcal{L}_{\mathfrak{w}}^* \)-formula

\[
K_{v_1} \exists y((\exists x/K_{v_1}, y)(K_{v_2}/K_{v_1}, y)Sxy). \quad (3)
\]

The slash indicates imperfect information as can be readily seen from the bifurcation of the operators in (2).

Semantically, as explained above the idea of applying hidden elements informally means that the verifying team of players \( V \) as indeed the falsifying team of players \( F \) do not know all the previous moves that have happened in the game. Thus in the above formulas the choice of the possible world for \( K_{v_1} \) is hidden from \( F \) choosing for \( K_{v_2} \).

5. Object identification and quantification

We already have gotten a glimpse of what is attained and what is still needed when quantifiers are added to epistemic logic. Firstly, let us clarify the concept of the model. A model \( M \) for \( \mathcal{L}_{\mathfrak{w}}^* \)-formulas is a tuple \( \langle \mathfrak{W}, \mathfrak{A} \rangle \), where \( \mathfrak{W} \) is a \( \tau \)-structure \( \langle \mathcal{W}, D_{\tau} \rangle \) of a signature \( \tau \) of a non-empty set of possible worlds \( \mathcal{W} \) and a non-empty world-relative domain \( D_{\tau} \) of individuals. \( \mathfrak{A} \) is an interpretation \( \langle \pi, R \rangle \), where \( \pi \) attaches to each possible world a \( \tau \)-structure together with a valuation from terms to the domain of the world. For example, \( S^{\pi(w)} \) means that the predicate \( S \) is interpreted in the world \( w \).

We then need a valuation \( g: X \rightarrow D_{\pi} \), which assigns individual variables in \( X \) to elements in \( D_{\pi} \). Let us extend this valuation to \( g^{\pi(w)}: \mathfrak{T} \rightarrow D_{\pi} \), which maps terms to the domains of individuals such that every \( n \)-ary function symbol \( f^n \) has a denotation of an \( n \)-ary (possible partial) operation in a domain \( D_{\pi} \), every variable \( x \in X \) has a denotation \( g^{\pi(w)}(x) = x^{\pi(w)} \), every constant symbol \( c \in C \) has a denotation \( g^{\pi(w)}(c) = c^{\pi(w)} \), and every \( m \)-ary predicate symbol \( S \) has a denotation of an \( m \)-ary predicate (or relation) in a domain \( D_{\pi} \). Every term \( t = f(t_1, \ldots, t_m) \) can now be recursively defined as \( g^{\pi(w)}(t) = f^{\pi(w)}(g^{\pi(w)}(t_1), \ldots, g^{\pi(w)}(t_m)) \). In addition to a relational \( \tau \)-structure, \( \pi \) now attaches to each possible world a valuation \( g^{\pi(w)}: \mathfrak{T} \rightarrow D_{\pi} \).
The key concept here is the following. We enrich concrete knowledge of objects had been left intact. the semantics with a finite number of identifying For example, such patients may not be able to concepts) in $M$, by extending the valuation $g$ to a (partial) mapping from worlds to individuals, that is, to $g: X \rightarrow D_w^w$, such that if $w \in \mathcal{W}$ and $g$ is an identifying function, then $g(w) \in D_w$. These functions can be read so that individuals have only local aspects or manifestations of themselves in any particular world (cf. the semantics given in Hintikka, 1969). Two such functions may also meet at some world, and then part company.

The interpretation of the equality sign $\equiv$ (the identifying functional) now is

$$\langle M, w_0, g \rangle \models x \equiv y \text{ iff for some } w_i, w_j \in \mathcal{W};$$

$$\exists f \exists h \text{ such that } f(w_i) = h(w_j).$$

That is, two individuals are identical if and only if there are world lines $f$ and $h$ that pick the same individuals in $w_i$ and in $w_j$. World lines can meet at some world but then pick different individuals in other worlds; the two-place identifying functional spells out when they meet. Individuals within a domain of a possible world are local and need to be cross-identified in order to become specific and in this sense global.

There are two main reasons why we ought to resort to identifying functions or world lines in the semantics of epistemic notions. Firstly, as will be argued in Section 7, one is virtually able to see the locations of specific focus amidst agent’s epistemic alternatives. Secondly, they can be drawn in various ways depending on how objects are identified, which has interesting connections with actual neuroscientific phenomena. Related to this is the fact that this method conceptualises the domains of epistemic alternatives in the sense that individuals can be conceived in multiple ways, or from multiple perspectives, in world-relative domains.

As to the second point, an illustration of the connections that obtain between the concept of identification in quantified epistemic logic and cognitive neuroscience is that some aphasic patients suffering from injuries in brain tissue sometimes lose their ability to identify objects in an abstract or categorical way, while retaining the ability to process information more concretely, as if the acquainted or concrete knowledge of objects had been left intact. For example, such patients may not be able to recognise various shades of red as examples of the same colour, yet no difficulty occurs in identifying them with indexical notions (such as ‘my red tie’ of a certain patient reported in Rosenfield, 1995, p. 28). Consequently, many descriptive terms are abstractions or generalisations that are easily lost. Sacks (1986, p. 12) reports that a patient identified a rose as ‘about six inches in length. . . . A convoluted red form with a linear green attachment’. However, this patient’s identification of a rose was completed after smelling it. In the subsequent situation, the same patient described a glove as ‘a continuous surface. . . . infolded on itself. It appears to have. . . . five outpouchings, if this is the word’ (Sacks, 1986, p. 13). The glove was finally identified when the patient put it on and tried to use it: ‘My god, it’s a glove!’

These are by no means the only connections where neuroscience and logical aspects of knowledge meet, and further relations are identified in Section 9.

Meanwhile, what will be argued in the following three sections is that in multi-agent systems communicating with each other a new type of knowledge will emerge, namely that of focused knowledge. The term is new with this paper. It means knowledge of two or more agents focused on some set of objects so that individual agents’ knowledge may differ in certain qualities. Apart from being directly relevant to knowledge representation tasks in multi-agent systems and processes, the existence of focused knowledge underscores the importance of the distinction between specific versus non-specific object identification that inevitably arises in quantified logics of knowledge, which vindicates the existence of various forms of object identification showing up in cognitive and neuroscientific experiments.

6. New types of knowledge emerge in multi-agent systems

The concept of knowledge is ubiquitous in multi-agent systems. Received epistemic logics are not sufficiently expressive, however, to capture knowledge in all multi-agent configurations. By means of a brief case study, this section demonstrates that new
forms of quantified epistemic logics are needed in order to attain these new types of knowledge. These logics are along the lines introduced in previous sections, and are associated with powerful game-theoretic tools in handling all kinds of knowledge constructions.

One way of putting forth the case of this section is to recall that classical logic is logic of perfect information transmission: each evaluation step in truth-clauses is revealed to the next level. The same of course holds for classical epistemic and other modal logics. But the assumption of perfect information is clearly inadequate for multi-agent systems, where information is often uncertain and hidden from other parties.

For instance, concurrent information processing manifests imperfect information. Besides epistemic logic, a logic of concurrent processing may involve branched organisations of quantifiers:

\[ M = \forall x \exists y \forall z \exists w \ \text{iff} \ \exists f \exists g \forall x \forall z \ S(x) \equiv (z) \]

It has been shown in de Alfaro, Henzinger, and Mang (2000) that synchronous single-step control modules with static and fixed typing gives rise to these kinds of quantifier structures. (Originally, partially ordered quantifiers were introduced in Henkin, 1961.) Their meaning is given by existentially quantified Skolem functions \( \exists f \) and \( \exists g \) that can be viewed as winning strategies in the correlated semantic game.

To understand knowledge in communicating multi-agent systems we in fact need very similar kinds of orderings. To see this, suppose that a process \( U_2 \) sends a message \( x \) to \( U_1 \). We ought to report this by saying that ‘\( U_2 \) knows what \( x \) is’, and ‘\( U_1 \) knows that \( it \) (the same message) has been sent’. (\( U_1 \) might knows this, say, because the communication channel is open.) This is already a rich situation involving all kinds of knowledge. However, the kind of knowledge involved in this two-agent system cannot be captured in ordinary first-order epistemic logic.

This is because what is involved in this situation is the representation of two clauses ‘\( U_2 \) knows what has been sent’ and ‘\( U_1 \) knows that something has been sent’. Traditionally, the former is related to the knowledge \textit{de re} and the latter to the knowledge \textit{de dicto}. What is not involved is the representation of the clause ‘\( U_1 \) knows that \( U_2 \) knows’, nor that of ‘\( U_2 \) knows that \( U_1 \) knows’. The decisive question here is, how do we combine these clauses? It is easy to see that attempts such as the following three all fail.

\[ \exists x K_{U_2} \text{Message}(x) \land K_{U_1} \exists y \text{Message}(y) \]  
\[ \exists y (K_{U_1} \text{Message}(x) \land K_{U_1} \exists y \text{Message}(y) \land x = y) \]  
\[ K_{U_1} \exists y (\text{Message}(x) \land K_{U_2} \text{Message}(y)). \]

Likewise, an attempt is doomed that tries to use two variables and distinguishes between a message whose content is known (say, by a predicate \textit{Content}(x)), and a message that has been sent (by a predicate \textit{Sent}(y)):

\[ \forall x \exists y ((K_{U_1} \text{Content}(x) \land x = y) \land K_{U_2} \text{Sent}(y)). \]

This does not work because now \( U_2 \) comes to know what has been sent, which is too strong an assertion.

What we need is information-hiding alias branching, concealing the information about the choices for some possible worlds. Formally, this means that the following logical representation is needed:

\[ \exists x K_{U_2} \text{Message}(x) \land K_{U_1} \exists y (\text{Message}(x) \land x = y). \]

Equivalently, this can be expressed by the \( \mathcal{L}^* \) formula

\[ K_{U_1} \exists y (\exists x/K_{U_1}, y)(K_{U_2}/K_{U_1}, y)(\text{Message}(x) \land x = y). \]

Whenever we have concurrent processing in quantified epistemic logics, and whenever there are two or more agents involved, a novel type of knowledge will rear its head. As noted, the ensuing notion of knowledge is focused knowledge, and it can naturally be generalised to multiple agents.
7. Games, focused knowledge and imperfect information

7.1. Refinement on rules

In order to understand quantified epistemic logic $\mathcal{L}^*_\text{epi}$ with imperfect information, we still need to know what the notion of focus means in sentences like (10). To attain this we need to refine our game rules. Firstly, we assume that when $K_i$ is in the scope of $\exists x$, or is found in different rows from that of $\exists x$ given the two-dimensional branching notation (and the game has reached $w$), the individual picked for $x$ by $V$ has to be defined and exist in all worlds accessible from the current one. This assumption is motivated by the fact that the course of the play reached at a certain point in the game is unknownst to $F$ choosing for $K_i$. This will lead to specific knowledge (de re) of individuals, which can be correlated with games of imperfect information in the sense of extensive games to be given below.

The other proviso is that when $\exists x$ is in the scope of $K_i$, the individual picked for $x$ has to be defined and exist in the world chosen for $K_i$. This, in turn, will lead to the notion of non-specific (de dicto) type of knowledge.

Let $\varphi$ be an $\mathcal{L}^*_\text{epi}$-formula and let $B$ be a set of modal operators and variables already occurred in the game $G(\varphi, w, M)$ when an expression of the form $(Q/B)$ is encountered. The game rule for the hidden information is now the following.

$$(G.Q/B): \text{ if } \varphi = (Q/B)\psi, \ Q \in \{\forall x, \ K_i\}, \text{ and the game has reached } w, \text{ then if } Q = \forall x, \text{ a member } F_i \in F \text{ chooses an individual from the domain } D_{w_i} \text{ of individuals, where } w_i \text{ is the world from which the world chosen for the first modal operator in } B \text{ departed. The next choice is in } G(\psi, w, M).$$

If $Q = K_i$, then a member $F_i \in F$ chooses $w_i \in W$ in the model $M$ independently of the choices made for the elements in $B$, and the next choice is in $G(\psi, w_i, M)$.

Likewise for the team $V$’s choices.

This rule can be deciphered by writing the game out in its extensive form. Then we will automatically have a bookkeeping system of derivational histories of the plays of the game that tracks previously chosen worlds as well as the values chosen for quantifiers and connectives. The notion of ‘choosing independently’ with reference to the choices of worlds can, but does not have to be, taken to mean uniformity of such choices (spelled out in the uniformity of players’ strategy functions, see below), but can also mean as in the case of choosing individuals that such worlds are picked which are accessible from the worlds found by backtracking to the history from which the first operator in the sequence $B$ departs. The task of choosing between these differing interpretations is left to the modeller to decide.\footnote{Since the rule $(G.Q/B)$ refers to the location ‘the world from which the world chosen for the first modal operator in $B$ departed’, it is worth noting how close we come to what are known in the trade as hybrid modal logics (Blackburn, 2000), where one in effect uses terms in the object language in order to refer to individual worlds.}

Similarly as for propositional extensions, the winning conventions in the game $G(\varphi, w, M)$ for $\mathcal{L}^*_\text{epi}$-formula $\varphi$ are given by the truth and falsity of its atomic formulas, and the truth and falsity of complex formulas are given by the existence of a winning strategy for $V$ and $F$. Notwithstanding this, the overall game-theoretic content of the semantics needs to be enhanced by the following couple of elaborations.

7.2. Extensive forms for logic

Independent modalities mean that the player choosing for $K_i$ is not informed about the choices made for $K_j (j \neq i)$ in the other rows of a branching quantifier prefix (or with reference to slashed constituents). This can be brought to life by taking hints from the theory of games. We shall apply a partitional information structure $(I_i)_{i \in N}$ in the corresponding extensive games, which partitions sequences of actions (histories) $h \in H$ into equivalence classes (information sets):

$$\{S_i^j | S_i^j \in (I_i)_{i \in N}, \ h \sim h' \in S_i^j, h, h' \in H\}.$$
tories within an equivalence class are not known to a player $i$ whose equivalence class is in question.

7.3. Restrictions on strategies

The above description defines an extensive form, but not yet an extensive game. To do this, payoff functions $u_i(h)$ are needed to associate a pair of truth-values in $\{1, -1\}$ to terminal histories $h \in H$. Strategies will then be functions $f_i: P^{-1}(\{|\rangle\}) \rightarrow A$ from histories where players move to sequences of actions in $A$. If $i$ is planning his decisions within the equivalence class $S_i'$, annotated for him or her, the strategies are further required to be uniform on indistinguishable histories $h, h' \in S_i'$. In that case $f_i(h) = f_i(h'), i \in N$.

These definitions lead at once to the following observation:

Tracing uniform strategies along the game histories reveals in which worlds the specific focus is located.

To see this, it suffices to correlate information sets of an extensive game associated with the formula (10) with world lines of a suitable model $M$.

This requirement on strategies agrees with one that those of common, increasing, decreasing and open domains assumptions.

Remarks

A couple of remarks are in order. Firstly, the notion of uniformity sets some inevitable constraints on allowable models. In particular, at any modal depth $d$ (defined in a standard way) there has to be the same number of departing worlds. In addition, if we assume that the players can observe the set of available choices, the uniformity of strategies requires that the departing worlds have to in fact coincide for all indistinguishable worlds. In quantified contexts, this means that the domains have to coincide for any two worlds at the same depth $d$ that are members of the same information set in the sense of extensive game representation. This leads to a new stratified domains restriction different from those of common, increasing, decreasing and open domains assumptions.

Secondly, we may take $K_i$’s in different rows either to refer to simultaneous accessible worlds from the current one, or to detached submodels of $M$. In the latter case we actually evaluate formulas in a sequence of actual worlds, that is, in $\langle M, (w_0, w_1, \ldots, w_0, g), g \rangle$, breaking the models into detached submodels. The designated worlds in each submodel are independent from each other by definition. It needs to be noted that in that case there may still be world-lines that span across detached submodels, which provides a suitable connection between them (Pietarinen, 2001).

Thirdly, in a natural sense imperfect information games are not sequential but ones where players move simultaneously. For if I am ignorant of the earlier move, I can as well make the choice before that move, or in concert with it. The sequential structure of these extensive games is just an artificial

\begin{align*}
f_{V_2}(w'_1, a) &= f_{V_2}(w'_2, a) \quad \text{and} \quad f_{V_2}(w'_1, b) = f_{V_2}(w'_2, b). \\
Likewise, \quad g_{V_2} \quad \text{is} \quad \{a, b\}-\text{uniform:} \\
g_{V_2}(w'_1, a) &= g_{V_2}(w'_2, a) \quad \text{and} \quad g_{V_2}(w'_1, b) = g_{V_2}(w'_2, b). \\
\end{align*}
way of depicting the actions of players in a super-

ceptually sequential format. The identity in (12), as an

8. Further evidence for non-linearity: intentional identity

Another situation where variants of focused

knowledge and a fortiori epistemic independence are

all pervading is found in natural language sentences

with what after their introduction in Geach (1967)

have been known as ones with intentional identities.

These sentences have resisted purely semantic for-

malisations and explications over the years. A var-

iant of them is the following:

\[ U_1 \text{ knows that a message has been sent and } U_2 \text{ knows}
\[ \text{that it (the same message) has been controlled.} \quad (11) \]

As shown in Pietarinen (2001), the logical force of

this sentence is seen to be the following:

\[
K_{U_1} \exists x (\text{Message}(x) \wedge \text{Sent}(x) \wedge \text{Message}(y) \wedge
\text{Control}(y) \wedge x = y).
\]

(12)

As is readily seen from this branching symbolisation,

we can adopt previous definitions and receive a

semantic analysis via game-theoretic semantics of in-

perfect information. The notion of focused knowl-

dge arises here as well, but unlike sentences such as

(9) and (10), we have here two instances of non-

specific knowledge.

Pietarinen (2001) explains further the purely

semantic behaviour of these sentences. Among other

things, they are important in understanding natural

language anaphora in propositional attitude contexts.

They have cognitive significance in the formation of

the concept of identity. The identity in (12), as an

identifying functional that spells out when two

functions coincide, relates functions that in possible-

world semantics are among the first-class citizens

representing individuals in the model. Obviously,

functional coincidence denotes a complex object that

depends not only on which objects are identical in

which possible worlds, but also on which objects are

not identical in some possible worlds. These identity

relations may in turn depend on the way objects are

taken to persist from one world to another, which is a

question concerning not only purely logical aspects

of the model but also agent’s cognitive repertoire in

the process of drawing these links.

In previous attempts to analyse these sentences, it

has been argued that their meaning cannot be

explicated without recourse to pragmatic notions

such as common ground between the two agents

\( U_1 \) and \( U_2 \). However, the symbolisation of (11) by (12)

shows that such claims are ill founded, and a purely

semantic explication of intentional identity is

possible.\(^5\)

The overall argument in Sections 3–8 concerning

knowledge in simple two-agent cases is calculated to

show that there are true applications for logics that in

a well motivated sense go beyond all received

epistemic languages. The argument can of course be

extended to \( n \) agents. In representing knowledge in

multi-agent systems one needs new cognitively

justified epistemic languages that capture the notion

of focus of two or more agents, and do justice to

bounded versions of knowledge, information and

rationality. However, classical logics based on per-

fect, uninterrupted flow of information do not give us

enough expressivity in order to achieve this, and

hence we need to devise new logics alongside the

semantic framework that can conveniently be based

on the theory of games. Such games put before us a

concrete idea about how objects are identified by

agents.

In the remaining sections we switch the perspec-
tive to that of empirical findings in cognitive neuro-

---

\(^4\)Typically, simultaneous moves in extensive games relate to

concurrency in the sense that there is concurrent action if every

history crosses all information sets (Bonanno, 1992). However,

richer notions of concurrent games have been developed for

computational and logical purposes in e.g. Abramsky and Mellies

(1999), de Alfaro and Henzinger (2000).

\(^5\)Common ground will nonetheless be a vital assumption that

holds on the level of semantic games, but not on that of agents.
science, assessing their impetus for logical theories of knowledge.

9. Aspects of cognitive neuroscience in a logical perspective

9.1. Awareness and the implicit/explicit distinction

Connections between logical aspects of propositional attitudes (e.g. knowledge and belief) and neuroscientific findings concerning knowledge, perception and so on are to date pretty much uncharted. For example, species of epistemic logic known as awareness logic (Fagin & Halpern, 1988), originally calculated to deal with problems of logical omniscience in epistemic logic, can be made to reflect many neurological phenomena and dysfunctions, including unaware information processing, implicit versus explicit knowing and memory, prosopagnosia, blindsight, numb-sense, and so on. Primarily for didactic reasons, in the remainder of this paper I will confine myself to the propositional level.

To begin with logics of awareness, an agent can be said to be aware of a proposition \( p \) just in case the agent knows that \( K_i(p \lor \lnot p) \), that is, \( i \) knows that the law of excluded middle holds. This has sometimes been explicated in the sense of situation semantics (Barwise & Perry, 1983), which aims at articulating that the agent is aware of the proposition \( p \) precisely in those situations that support or preserve the truth-value of \( p \). Likewise, the agent is said to be unaware of \( p \) in situations that do not support or refute \( p \). In alternative terminology, in these situations there are no partial interpretations for sentences of which the agent is aware, that is, agent is not aware of propositions that fall into a truth-value gap or have the truth-value of Undefined (see Section 14).

Formally, awareness itself is typically a syntactic operator conjoined to an implicit knowledge operator, calculated to turn implicit knowledge of a proposition into explicit knowledge (Fagin & Halpern, 1988). That is, given implicit knowledge, which we denote by \( K_i \varphi \), it is transformed into explicit knowledge by defining \( K_i \varphi := K_i \varphi \land A_i \varphi \). The operator \( A_i \) attached to the proposition \( \varphi \) means that ‘\( i \) is aware of \( \varphi \)’. It is explicated in Fagin and Halpern (1988) by syntactic means of denoting an ‘aware’ or ‘attended’ proposition. To this effect, a mapping \( \mathcal{A}(w) \) is used, which takes a world \( w \) to propositions, namely to the set of propositions of which \( i \) is aware.

This proposal has to be supplemented with some remarks. Firstly, the suggestion that the awareness operator can be interpreted in different ways depending on the purpose at hand, is not motivated by conceptual or cognitive concerns, but by computational ones. But computation is an unlikely candidate for a general theory of cognition. In Fagin and Halpern (1988) such readings are proposed as ‘an agent is aware of \( p \) if she is aware of all concepts that it contains’, or ‘\( p \)’s truth-value can be computed within an interval \( T \)’. The snag in these is the danger of reducing the whole system into syntactically characterisable sets of propositions intersected with sets of those propositions falling within the agents’ range of attention. This explication leaves little room for other, possibly more versatile and useful semantic, conceptual and cognitive investigations related to different kinds of concepts of awareness.

Rather than addressing this conceptual and motivational problem, our contribution here is confined to an investigation of a couple of logical variations to which the notion of awareness gives rise. In particular, new evidence is derived from cognitive neuroscience that can be reflected in logical and semantic frameworks. As will be seen, various experiments in neuroscience have demonstrated that implicit aspects of knowing, remembering and perception are fruitful fields of research for the multidisciplinary study of cognition and awareness (Dienes & Perner, 1999; Underwood, 1996; Weiskrantz, 1997; Wilson, Clare, Young, & Hodges, 1997). This suggests that in the corresponding corners of logic dealing with, for instance, propositional attitudes of knowledge, memory and perception, we are well advised to try to develop mechanisms that transform implicit propositional attitudes into fully recognised, explicit ones falling within agent’s range of aware or conscious attitudes.

9.2. Blindsight

Certain curious damage to the visual cortex, dubbed blindsight in neuroscientific literature, has aroused considerable literature among neurologists
and cognitive scientists (Weiskrantz, 1997). The key finding has been that patients may lose phenomenal seeing in the contralateral half of the visual field while retaining residual capacity. The residual capacity is something the patient is not aware of, at least not in the same sense as he or she is aware of objects and things within the parts of the field of phenomenal vision. This supports the distinction between implicit and explicit modes of information processing, or logically speaking between aware and unaware propositions, given a suitable propositional attitude operator attached to the propositions representing direct objects of seeing.

In order to accommodate this basic phenomenon into the logical system of awareness, we need to take into account what the neural dysfunction of blindsight suggests, namely that given a possible world w, agent i implicitly sees φ (i.e. Sφ), if and only if (M, w) |= φ for all w′ ∈ [w]i such that φ ∈ ωi(w). The set of propositions ωi(w) of which i is aware at w is interpreted as being constructed from those propositions ψ of which i explicitly sees that φ.

Supported by actual neuropsychological findings, by means of this apparatus and the distinction between two modes of seeing, the logic of perception (e.g. the one originally formulated in Hintikka, 1969) can be enhanced by new kinds of logical rules. Among them is one saying that a patient explicitly sees φ, but nonetheless does not see some subformula ψ of φ. Let Sub(φ) mean a set of subformulas of φ. Explicit perception can now be captured by excluding the rule ‘If φ ∈ ωi(w) and ψ ∈ Sub(φ), then ψ ∈ ωi(w)’ from the set of formation rules for the logic of implicit and explicit perception. In words, this reflects the fact that some contralateral subformulas may be severed from the formulas within the set of explicitly seen propositions. It is readily seen that this is what transpires in blindsight, where patient’s overall visual scenery may consist of parts that register stimuli only in some roundabout way, without any conscious visual experience of it.

9.3. Unilateral neglect

There are further examples in cognitive neuroscience echoing the distinction between implicit and explicit aspects of attitudes. In unilateral neglect, patients’ visual range of attention is focused only on left or on right visual field, and their behaviour constrained accordingly. Yet, experiments have demonstrated some conceptual dependencies between the two fields, such as identity relations, leading to a preference toward situations presented to one half of the visual field over the other. This suggests that visually severed information is after all processed by the brain. The experiments on unilateral neglect are not limited to visual perception, and take auditory and haptic stimuli into consideration too.

Implicit aspects of this kind of recognition can logically speaking be expressed by omitting the law of commutativity from the set of formation rules of the logic of implicit and explicit perception, that is, by omitting ‘(φ ∧ ψ) ∈ ωi(w) if and only if (ψ ∧ φ) ∈ ωi(w)’. The absence of this rule would make the logic to invalidate cases like ‘if S(A, φ ∧ ψ) then S((ψ ∧ φ) ∧ A)’, where φ can be taken to correspond to, say, subject i’s left visual input, and ψ can be taken to correspond to his or her right visual input.

9.4. Prosopagnosia

Some other, closely related implicit aspects of information processing in the brain come to the fore in prosopagnosia, an incapacity of overtly recognizing and pointing out familiar faces. Yet, the autonomic nervous system may signal differences between familiar and unfamiliar faces. In order to properly investigate the implications of this phenomenon to the logic of awareness, we would actually need to extend the propositional epistemic logics of awareness to cover also quantified cases. This is because in analysing prosopagnosia, we would need to be able to speak about individuals and choose them as values of quantified variables, and then go on to formulate statements such as ‘an agent knows who someone is’. Accordingly, locutions of this kind would then need to admit both explicit and implicit readings. We refrain from presenting such extensions here.

10. Aspects of awareness and memory

Let us then touch on some of the emerging issues in the relation between neuroscience and logical
aspects of memory. Firstly, it needs to be remarked that given the variability of all kinds of deficits that memory systems have been observed to cause in patients, we seem to have a much more heterogeneous concept at hand than when dealing with perception. In general amnesia, all kinds of distracting phenomena can occur. To limit the discussion here to just one aspect of it, we note that it has been shown that in priming experiments amnesics exhibit impaired conscious access to memories. While not exposing overt recall, subjects can identify things in a series of tests despite their insistence of not recalling any of the past tests (Schacter, 1987).

Furthermore, in covert, imperfect recall, the kind of awareness exhibited is both conceptually and neurologically different from that of perceptual awareness. It results from an impairment of those mechanisms that are responsible for limiting conscious access to memories. On the other hand, dysfunctions in perceptual awareness result from impairments to sensory systems.

Yet, the kind of information processing in the brain that relates to memory is seen to conveniently divide into implicit and explicit halves. For example, in Schacter (1987) it is argued that implicit aspects of memory do not involve conscious or intentional recollection of experiences, while explicit memory does involve such conscious recollection. Unlike claims to the effect that implicit memory may simply lack suitable context in order to be explicitly recollected (e.g. spatio-temporal, semantic, emotional etc.), these claims have remained without comparable experimental backing.

Therefore, as strongly suggested by many experiments, if some kind of implicit memory actually exists then the ensuing distinction between implicit and explicit memories gives rise to interesting logical properties. First of all, we would need to distinguish between explicit memory or explicit recall (retrieval) denoted by $R \varphi$, and implicit memory or implicit recall, denoted by $I \varphi$. The relation between them is given by $R \varphi := I \varphi \land A \varphi$, where $\varphi \in \mathcal{A}(w)$.

But if so, then iterations of remembering something, as indeed in knowing something in general, pertain to explicit, not implicit memory. That is, if $R \varphi$ then $KR \varphi$ (i.e. one knows that one remembers $\varphi$), and if $R \varphi$ then also $AR \varphi$ (i.e. one is aware of one’s memory of $\varphi$). But the property of transitivity of frames in the possible-worlds structure (i.e. if $w_1 \in [w_0]_R$ and $w_2 \in [w_1]_R$ then $w_2 \in [w_0]_R$) is dubious in explicit remembering, and hence the principle that says that whenever $R \varphi$ then $RR \varphi$ is likely to fail. It is not plausible to implicitly remember (know, see etc.) to implicitly remember something, and it is not plausible to explicitly remember to implicitly remember something, and it may not even be plausible to implicitly remember to explicitly remember something.

Consequently, the status of iterated attitudes and the positive and negative introspection properties has to be re-examined especially with respect to explicit attitudes, irrespectively of whether they pertain to one’s memory, knowledge or perception.

These properties and the general distinction between implicit and explicit memory can then be embedded into a kind of logic of memory and the associated possible-worlds semantics expounded, for example, in Aho and Niiniluoto (1990).

11. Different senses of awareness

We would still like to know what the source of these different kinds of manifestations of awareness is. Is there some underlying general mechanism in information processing responsible for them, or do they illustrate transparently different phenomena?

This question cannot be answered in any fully satisfying way given the confines of this paper, but one important related aspect can nonetheless be put forward. There does not seem to be any invariable threshold that demarcates between those stimuli (propositions) subjects are aware of and those they are not aware of. At the very least, no general mechanism seems to be available in the human brain that could function as a cut-off point for awareness. Rather, awareness is something that comes in all shades and grades. It is even conceivable that there be objects or propositions in the subject’s grasp of which she is not aware neither unaware, a kind of agnostic cogitation. This distorts the characterisation of syntactic awareness as knowledge of the law of excluded middle. The lack of a threshold is nonetheless something that can be reflected in the logic of explicit and implicit processing by keeping the
interpretation of the set of propositions $A_i(w)$ free from any such presuppositions (see the next section).

At this point, one may also raise a question of whether there exists also a third mode of awareness, a kind of reflective awareness of oneself or his or her physiology. Indeed, this mode can be readily distinguished from the previous two, perceptual awareness and awareness or range of attention pertaining to memory in the sense of being able to retrieve traces of one’s past. Psychophysicists have even given a name to awareness of the third kind: proprioception. However, there is precious little it seems to the extent in which logical considerations can be brought to bear on this admittedly intriguing cognitive and psychophysiological phenomenon. We are not dealing with the same conceptual entity as given by the notion of self-identification, for example. For self-identification contributes at least modestly to the interplay between logic and cognitive neuroscience. (Self-identification is sometimes called \emph{de se} knowledge in philosophical literature, see Lewis, 1979.) By contrast, proprioception has to do with the idea of self-monitoring, although self-monitoring is a wider phenomenon. Proprioceptive monitoring refers to the capacity of recognising one’s current bodily state and position in his or her environment, and to the capacity of properly reflecting on those states and positions. As described in Sacks (1986), its malfunctions can cause curious effects in patients’ behaviour, but they do not appear to result from any deficits in what is reported by one’s propositional attitudes.

The different senses of awareness have not gone unnoticed in the neuroscientific and neuropsychological literature. Weiskrantz (1997) distinguishes cases where patients are properly reported to be ‘aware of something’ from those where they are reported to be ‘aware that something holds’. In blindsight, for example, patients may be unaware of the very material being presented, whereas in amnesic syndrome, they are unaware that a positive response in the test demonstrates the intactness of a capacity (Weiskrantz, 1997, p. 45). This is reminiscent of the traditional philosophical distinction between attitudes \emph{de re} (i.e. ‘of a thing’, or specific attitude) and attitudes \emph{de dicto} (i.e. ‘of saying’, or non-specific attitude) in the sense where the former type of awareness can be described as ‘specific awareness \emph{de re}’ and the latter ‘non-specific awareness \emph{de dicto}’.

The distinction between these two types of attitudes resorts to quantification over individuals. It would be productive to try to bring out this distinction in one of the predicate extensions of the relevant epistemic logic of awareness. In such a framework, awareness as a syntactic filter would no longer be viable but needs to be interpreted as a proper modal operator. Rather than undertaking this way of extending the language here, in the following two sections we turn to another formal aspect of awareness, namely that of its general and dynamic character.

12. A model for general awareness

Awareness is a vague notion, and this is a fact that comes out clearly in the difficulties of precisely knowing when something falls within the reach of awareness and when something falls outside it. One thus needs a logic that is capable of reflecting this dynamics. But here one runs into several complications when trying to represent sets of agent’s knowledge, belief or perceptual propositions by a static model provided by the traditional possible-worlds semantics. For one of the remarkable characteristics of cognition is its capacity of directing the focus of attention rapidly from one target to another, without any feeling of interference by conscious, volitional instructions. But when attention shifts take place, it is no longer the case that precisely the same logical principles will hold from one range of focus of attention to another. We need to enrich the language.

The suggestion that can be made here is to consider the awareness filter not as a singular operation but as a family of different filters, each one tailored for different purposes. Let $C_i(w)$ be a non-empty set of subsets of the worlds $\mathcal{W}$. Such a set of sets of worlds will be denoted by $\{T_1, \ldots, T_n\}$. Hence, given a world $w$, $i$ would consider any $T_j$ possible, given some further contextual or cognitive parameters. Now, we can define the awareness operator for each $T_j$ of these frames of minds. That is, let $C_i(w)$ be a function $C_i: \mathcal{W} \to 2^{2^m}$ for each $i$, and let $\mathcal{A}'$ be a family of functions consisting of sets of awareness...
operators $\mathcal{A}'(w)$, mapping each $w$ to a set of propositions, given a set of worlds $T_j \in C$. The advantage is that each $\mathcal{A}'(w)$ of $\mathcal{A}'$ can be given some desired strictures of its own, for instance in accordance with the examples given in previous sections. This generalises the approach presented in (Fagin et al., 1995, pp. 342–346).

The clauses for knowledge and awareness would be thus:

\[
\langle M, w \rangle \vdash K_i \varphi \text{ iff for some } T_j \subseteq C_i(w)
\]

and for all $w' \in T_j$, $\langle M, w' \rangle \vdash \varphi$.

\[
\langle M, w \rangle \vdash A_i \varphi \text{ iff } \varphi \in T_j \subseteq C_i(w) \text{ for all } 1 \leq j \leq n.
\]

That is, one has to specify which awareness operation one is speaking about.

In general, this approach would be particularly useful in modelling aspects of belief instead of knowledge. One reason is that beliefs of a single agent can often be inconsistent. By means of the family of awareness filters, we cannot only model inconsistent explicit belief, but also see what kind of effects the fact of agent being aware of such inconsistencies may generate. For instance, let $B_i$ be a modal operator for explicit belief. Now if $B_i \varphi \land \neg B_i \varphi$ and $A_i'(B_i \varphi \land \neg B_i \varphi)$, $i$ would strive to switch into a new frame of mind that is consistent with the limitations given in it. In other words, $B_i \varphi \land \neg B_i \varphi$ and $A_i'(B_i \varphi \land \neg B_i \varphi)$ together imply $A_i'(\neg(B_i \varphi \land \neg B_i \varphi))$, $j \neq 1$.

One consequence of generalised epistemic logic of awareness from the point of view of doxastic notions is that inconsistent sets of beliefs—e.g. those argued to arise in relation to so-called puzzles of beliefs (Kripke, 1979)—can be represented and analysed in the resulting logic of general awareness for beliefs. We skip this issue here and go on to address another insight to doxastic attitudes in the next section.

13. Towards a generalised doxastic logic

Since logics of belief (doxastic logics) are also epistemic logics, in the light of the previous discussion they receive a particularly interesting semantic treatment. First of all, let us note that the logic of awareness communities can be given an interesting game-theoretic twist. As above, let us partition the doxastic possible-worlds structure into a nonempty set of subsets of $C_i(w)$. Each of these subsets represents agent’s frame of mind or a local belief set, where the actual choice of the local set $T_j \subseteq C_i(w)$ depends on factors like agent’s current focus, timing, the range of attention, progress of learning, or the like (cf. Fagin et al., 1995).

To treat these operators and the ensuing languages game-theoretically, one defines the appropriate game rules as follows. $(B_i^d \varphi$ is the dual of the belief operator $B_i \varphi$, defined by $\neg B_i \neg \varphi)$.

\[
(G, B_i \varphi): \text{ if the game has reached the world } w \in W \text{ and the sentence } \varphi, \text{ then the verifier } V \text{ chooses } T_j \subseteq C_i(w) \text{ where } T_j = \{w_1, \ldots, w_n\}, \text{ whereupon the falsifier } F \text{ chooses } w' \in T_j.
\]

\[
(G, B_i^d \varphi): \text{ if the game has reached the world } w \in W \text{ and the sentence } \varphi, \text{ then the falsifier } F \text{ chooses } T_j \subseteq C_i(w) \text{ where } T_j = \{w_1, \ldots, w_n\}, \text{ whereupon the verifier } V \text{ chooses } w' \in T_j.
\]

These are two-step rules. This is also reflected in the winning conventions, which say that if for all $w' \in T_j$, $p$ is true in $M$ at $w'$ according to the given valuation $g$ then $V$ wins, and if there exists $w' \in T_j$ such that $p$ is false in $M$ at $w'$ according to the given valuation $g$ then $F$ wins. Dually, if for all $w' \in T_j$, $p$ is true in $M$ at $w'$ then $F$ wins, and if there exists $w' \in T_j$ such that $p$ is true in $M$ at $w'$ then $V$ wins. As usual, the formula $\varphi$ of the language of generalised doxastic logic is true in $M$ at $w$, if and only if there exists a winning strategy for the player who initiated the game as $V$, and false in $M$ at $w$, if and only if there exists a winning strategy for the player who initiated the game as $F$.

This approach presents us with a useful new logic of belief with increased cognitive justification in its capacity of representing contradictory beliefs even more generally than is possible by the general awareness logic of the previous section.

We can put these expressions in plain words via a temporal explication. What $B_i \varphi$ can be taken to mean is that ‘$i$ sometimes believes $\varphi$’, while its dual $B_i^d \varphi$ says something like ‘$i$ always believes $\varphi$’, or ‘$i$ believes $\varphi$ all the time’. This is by no means the only colloquial explanation of these notions imaginable, however. An even more general perspective is that by localising agent’s beliefs to several mind-frames.
we can obtain generalised notions of belief, including intermediate locutions like ‘i mostly believes that’, ‘i seldom believes’, or ‘i almost always believes that’. There are innumerable further adverbial modifiers that can be envisaged. By continuing in this manner, a novel interpretation of modalities will be obtained analogous to that of a generalised notion of quantifiers (Barwise & Cooper, 1981), albeit couched in a game-theoretic setting. We note that there can be game rules which say that instead of choosing \( T_j \), a player chooses \( X \in T_j \), where \( X \) is adjoined with a generalised quantifier expression such as \textit{Most}, \textit{Few}, \textit{Many}, and the like. A comprehensive investigation of this direction remains to be carried out, but I believe that the ensuing generalised modalities will be of utmost importance for a variety of inquiries.

14. Partiality and games in epistemic logic

The previous suggestions towards a generalised notion of a belief readily point at another cognitively oriented logical issue. For what was done was a separation of truth and falsity for belief sentences from each other, which in general gives rise to logics that are partial. On the other side, vague or partial concepts in human cognition, such as perception and identification, are widely debated not only in the philosophy of mind and language but also in cognitive science. The term that often surfaces here is that of ‘fuzziness’ of our everyday concepts. Yet at least for logical purposes, it often suffices to have logic whose truth is defined so that there will be truth-value gaps, rather than find refuge in any full-blown fuzzy logic, whose explanatory value as regards to the motivations for having vagueness is close to zero.

As soon as we have truth-value gaps at our disposal, we are ipso facto dealing with the possibility of controlling and regulating semantic information flow within formulas of logic. For the truth (respectively, falsity) of sentences is in the game-theoretic semantics defined as an existence of a winning strategy for the verifier (respectively, the falsifier). But if a winning strategy does not exist for one of the players, it does not follow that the adversary automatically has a winning strategy. Hence the logic in question becomes partial. It is vital to add that the language can, but does not have to be, completely interpreted: the truth-value gaps in the case of complete interpretation would arise at the level of complex formulas. There are thus several routes to partiality.

To see the relation between partiality and the game-theoretic interpretation of epistemic logic, let us observe that a partial model is a triple \( M = (\mathcal{W}, R, g) \), where \( g' \) is a partial valuation function \( g': \mathcal{W} \to (\Phi \to \{\text{True}, \text{False}\}) \), assigning to some proposition letters in \( \Phi \) a subset \( g'(\Phi) \) of a set of possible worlds \( \mathcal{W} = \{w_0, \ldots, w_n\} \) for which \( p \) is true. The relation \( \models^+ \) then means positive logical consequence (i.e. formula being true in a model), and \( \models^− \) means negative logical consequence (i.e. formula being false in a model):

\[
(M, w) \models^+ K \varphi \iff (M, w') \models^+ \varphi \text{ for all } w' \in \mathcal{W}, w' \in [w]_{\rho},
\]

\[
(M, w) \models^− K \varphi \iff (M, w') \models^− \varphi \text{ for some } w' \in \mathcal{W}, w' \in [w]_{\rho},
\]

Similarly for the dual modality \( \neg K' \neg \varphi \).

However, an illuminating alternative way of looking at the phenomenon of partiality in epistemic logic is by games that are ones of imperfect information rather than perfect information. As seen above, such games can be applied to both propositional and first-order systems. This time partiality arises at the level of complex formulas, and does not need partial models with partial valuation and hence incomplete interpretation. This is because partiality is simply a result of assumptions concerning information transmission between the players of the semantic game.

This understanding of partiality may seem to have little to do with human cognitive limitations as such, as it is a logical fact about the transmission of information. However, if besides partiality at the level of complex sentences, atomic formulas are partially interpreted, one is reflecting various limitations of knowledge in the sense of failing to assign interpretations to all atomic formulas. It is of course perfectly feasible to combine both of these two senses of partiality.

As such, the game-theoretic framework advocated
in this paper is not compulsory for partiality, although it is instructive to see how it can be made to account for it, as an alternative to the more customary truth-conditional semantics. For instance, the truth-conditions for the generalised doxastic logic expounded in the previous section can be rewritten as follows:

\[ \langle M, w \rangle \models \uparrow B_i \varphi \iff \text{for some } T_j \in C_i(w) \]
\[ \text{and for all } w' \in T_j, \langle M, w' \rangle \models \uparrow \varphi. \]

\[ \langle M, w \rangle \models \downarrow B_i \varphi \iff \text{for some } T_j \in C_i(w) \]
\[ \text{there exists } w' \in T_j, \text{ s.t. } \langle M, w' \rangle \models \downarrow \varphi. \]

\[ \langle M, w \rangle \models \uparrow B_{ij} \varphi \iff \text{for all } T_j \in C_i(w) \]
\[ \text{there exists } w' \in T_j, \text{ s.t. } \langle M, w' \rangle \models \uparrow \varphi. \]

\[ \langle M, w \rangle \models \downarrow B_{ij} \varphi \iff \text{for all } T_j \in C_i(w) \]
\[ \text{and for all } w' \in T_j, \langle M, w' \rangle \models \downarrow \varphi. \]

What is essential in partiality is that the falsification conditions are separated from the verification conditions.

Furthermore, in addition to the two senses of partiality (partial interpretations versus partiality of complex formulas), we can also entertain with ‘non-standard’ notions of partiality. The following clauses illustrate how some further variations can come about in complex formulas:

\[ \langle M, w \rangle \models \uparrow K_i^\circ \varphi \iff \]
\[ \text{for all } w' \in W, w' \in [w]_{\mu_i}. \]

\[ \langle M, w \rangle \models \downarrow K_i^\circ \varphi \iff \]
\[ \text{not } \langle M, w' \rangle \models \uparrow \varphi \text{ for some } w' \in W, w' \in [w]_{\mu_i}. \]

\[ \langle M, w \rangle \models \uparrow K_{ij}^\circ \varphi \iff \]
\[ \text{not } \langle M, w' \rangle \models \downarrow \varphi \text{ for all } w' \in W, w' \in [w]_{\mu_i}. \]

\[ \langle M, w \rangle \models \downarrow K_{ij}^\circ \varphi \iff \]
\[ \text{for some } w' \in W, w' \in [w]_{\mu_i}. \]

The formula \( K_i^\circ \varphi \) captures the idea that the epistemic sentence is true in \( \langle M, w \rangle \) precisely when \( \varphi \) is true in all \( i \)-accessible worlds from \( w \), and false in \( \langle M, w \rangle \) when \( \varphi \) is not true in some \( i \)-accessible world from \( w \). The formula \( K_{ij}^\circ \varphi \), on the other hand, says that the epistemic sentence is true in \( \langle M, w \rangle \) precisely when \( \varphi \) is not false in every \( i \)-accessible world from \( w \), and false in \( \langle M, w \rangle \) when \( \varphi \) is false in some \( i \)-accessible world from \( w \). The standard interpreta-

tion clearly subsumes the truth-conditions for \( K_i^\circ \varphi \) and the falsity-conditions for \( K_{ij}^\circ \varphi \). The duals \( L_i^\circ \varphi \) and \( L_{ij}^\circ \varphi \) are defined accordingly.

Games for non-standard clauses change the rules for winning conditions to weaker ones. For \( \# \)-modalities we have:

\( \text{(G.atom\#)} \): if \( \varphi \) is atomic, the game terminates. \( V \) wins if \( \varphi \) is true, and \( F \) wins if \( \varphi \) is false.

Besides winning conventions, also the definition of winning strategies can be partialised to non-standard ones by stipulating that the existence of a winning strategy for the verifier means non-falsity. Likewise, the existence of a winning strategy for the falsifier can be taken to mean non-truth. This gives rise to players that are to be viewed as non-falsifiers and non-validators, not verifiers and falsifiers.

There are neuroscientific motivations for having partiality in logic, briefly discussed in the next section. The question of cognitive justification of these latter kinds of non-standard clauses needs to be left for further occasions, however.

15. Experimental evidence for the implicit/explicit distinction and awareness

A large amount of experimental work exists on neuropsychological deficits. They have for long vindicated that human information processing and knowledge bifurcates into explicit and implicit halves. These terms are notoriously vague, but as noted, the basic idea is that implicitness is a property of information that is manifested on performances that do not require conscious access or recollection of information or experiences. By contrast, explicitness does require one to be conscious of the information possessed. In the light of logic, this distinction was in previous sections reflected in the concept of awareness coming together with the propositional approach to knowledge.\(^6\)

\(^6\)A caveat is that the use of the notion of awareness in understanding consciousness or conscious access has its pitfalls, see e.g. Revonsuo & Kamppinen (1994), Young (1994) and Weiskrantz (1997).
In neuroscientific studies on vision, a distinction is customarily drawn between two functional systems, one being responsible for object perception and the other for perception between the relations of objects (Milner & Goodale, 1995; Underwood, 1996; Ungerleider & Mishkin, 1982; Vaina, 1987). A lively discussion exists on whether this distinction is to be made on the basis of different functional roles of the two cortical pathways or on the basis of their neurological structure. In the light of recent evidence, the neurological structure seems to be more complicated (Milner & Goodale, 1995). Already Van Essen and Gallant (1984) presented data on visual processing by identifying the visual cortex as a complex hierarchy of brain areas that have their own functional specialisation while retaining complex patterns of connectivity between them. This reflects the diversity of visual tasks, including the role of motion in visual perception.

At the very least, there seems to be a good deal of interaction between the two systems. Emphasised by Bechtel, Mandik, Mundale, and Stufflebeam (2001), the idea of two visual streams plays an important integrative role in the theories of visual processing. The distinction has its counterpart in the epistemic logic of quantified notions (Hintikka, 1989), where one is indeed equipped to make logically meaningful distinctions between the statements that require identification of objects on the basis of their public identity (‘I know who the winner of the race is’) versus their spatio-temporal, indexical identity (‘I know that the player over there is the winner’).

As has been aptly put by Dienes and Perner (1999), these two visual information processing systems identified with different neurophysiological pathways (dorsal and ventral) can be characterised by saying that the information in the dorsal system (the ‘what’ system) is unconscious, not used for localisations about external things, reliable in the sense of not producing illusions, used for action, and of limited duration. On the other hand, the information in the ventral path (the ‘where’ system) is conscious, illusory, about statements concerning perception, and used for action after some delay. But if the distinction is put in these terms, one sees that the distinction between implicit information (as represented by the ‘what’ system) and explicit information (as represented by the ‘where’ system) is applicable. One does not necessarily need to go beyond the propositional level of epistemic logic in tackling important neurological distinctions.

Perceptual identification is also related to experiments that test implicit versus explicit memory and recollection. This distinction has often been assimilated with that of direct vs. indirect tests, as direct tests appear to require subjects to be aware of the stimulus presented to them (Reingold & Merikle, 1988; Richardson-Klavehn, Gardiner, & Java, 1996). This is not the level of explanation that would tell implicit memory apart from explicit memory in terms of its logical representation. Better evidence comes from studies of brain-damaged patients with organic amnesia, who often show skills that appear to require memory or recollection of facts of which they nonetheless remain unaware (Schacter, Bowers, & Booker, 1989; Squire, 1992). In fact, there are various forms of implicit memory (Church & Schacter, 1994; Graf & Ryan, 1990; Lewandowsky & Murdock, 1989; Schacter & Tulving, 1994), providing experimental vindication of the kind of logic of memory presented above (Section 10) where the awareness filter that dissociates implicit from explicitly recollected facts may vary according to the logical laws permitted by the system.

Karmiloff-Smith (1992) argues that conscious access to data is acquired in a stepwise fashion during human cognitive development, and that there seems to be no definite breakpoint in learning where implicit conception of information would turn into an explicit conception. If we want to make a good logical sense of the general idea of the dynamics of the implicit/explicit distinction, we would be well advised to adjoin our propositional logic of knowledge and other attitudes with the notion of vagueness. This does not imply that the distinction between explicit and implicit would lose its importance: in logics of vagueness borderline cases are inevitable, which only means that some of the logical laws such as the law of excluded middle (and perhaps the law of contradiction) are no longer valid. But if so, then the approach to partiality presented in Section 14 is in fact the needed proper emendation to the received epistemic logic of awareness.

A further reflection of this is found in the logic of awareness communities (Section 9), where a finite number of different filters may exist that turn implicit knowledge, memory or some other attitude into an explicit one. But if we want to definitely
address the problem of vagueness, then a further partialisation is desirable.

In a similar vein, there is the multiple memory systems view put forward by Schacter (1987), which maintains that implicit and explicit aspects of memory are reflections of the operation of separate subsystems in the brain, characterised by different rules and performance. But if so, then precisely the kind of logical separation of remembering into two halves is vindicated. Even further, separate subsystems entertaining separate and even contradictory rules can be attained by the general awareness model of attitudes of Section 12.

Many studies in cognitive science and cognitive neuroscience have also been devoted to the issue of implicit versus explicit learning processes in skill acquisition (Slusarz & Sun, 2001) and in artificial grammar learning (Dienes, Altmann, Kwan, & Goode, 1995). It remains to be seen what their overall impact to logical and semantic theories will be.

16. Conclusion

What do epistemic logic and cognitive science have to do with each other? A number of empirical findings make a case in favour of the more expressive, somewhat radically revised conception of logic of propositional attitudes of knowledge and belief. Some of these findings pertain to notions of cognitive science, some to the experimental side of cognitive neuroscience and neurophysiology, while some are logical and semantic in nature.

To start with the semantic and logical motivation, what is observed is that the dynamic method of games in the semantics of epistemic notions is useful in attaching semantic attributes to formulas with various propositional attitudes, in particular when these formulas need to capture non-compositional forms of information hiding. Despite many of the systems described here being propositional, the addition of quantifiers would in the end be indispensable, since full notions of knowledge do not exist independently of how objects are identified, which in turn is intimately connected with the question of how various flows of information are controlled in epistemic logic, and how we actually individuate objects.

One consequence of the non-linear and non-compositional nature of the logical systems is that the game-theoretic stance gives rise to partiality at the level of complex formulas, dispensing with partially interpreted models and partial interpretations of atomic formulas. Partiality and the truth-value gaps are results of entirely classical assumptions concerning information transmission between the players in particular semantic games. This in turn implies that one is able to generalise the typical semantic approach to partial modal logics presented, for instance, in Jaspars and Thijsse (1996), which are based on the concept of partial models. There are also additional cognitive and experimental motivations for being able to produce such logics that have truth-value gaps given, among other things, by the overall vagueness of the notion of the implicit/explicit distinction arising in cognitive neuroscience.

Far from remaining on the classical level of a static universe of possibilities, the logical aspects of awareness are seen to implement many actual cognitive notions such as implicit versus explicit aspects of awareness, and the existence of a multiplicity of awareness filters within a single mind.

A wider outcome is that with the aid of game-theoretic semantics, possible-worlds structures are limbered up into a more dynamic mould than would be permitted by the classical truth-conditions. Just to mention one course of further research: even the interpretations of atomic formulas (game-theoretically, the values of the payoff functions) can be made to vary and hence multiplied according to the histories that have been traversed in order to arrive at them. This ties the semantics of the system more closely within the kind of logical approaches to memory suggested above.

On the cognitive and experimental front, perhaps the most stimulating result for the purposes of this paper is the demonstration by a couple of examples of the kind of interplay that the logical notions of a range of propositional attitudes, including knowledge, belief, perception and memory, entertain with some of the recent findings in cognitive neuroscience and neuropsychology. For just one thing, these conceptual issues raise their heads in systems that are controlled by languages involving epistemic notions,
for such systems will inevitably need some naming relations instead of generic terms in order to fulfill the tasks they were set out to do. These neuroscientific results may not revolutionise the semantics of propositional attitudes or multi-agent systems whose knowledge is intended to be represented by them, but they should not be ignored as they offer privileged and hitherto by and large unexplored insights into logical and multi-agent systems. Given the current proliferation of experimental research and investment in cognitive neuroscience and studies on awareness and consciousness, the interplay between logic and cognition is likely to reach increasingly wider and become increasingly prominent, encouraging fresh perspectives both from logical and semantic fields and from cognitive and neuroscientific communities.

By contrast to the suggestions in the literature to the effect that the awareness operators should be interpreted in different ways depending on the modelling purposes—suggestions that are not really motivated by actual conceptual or cognitive concerns but merely by computational ones—their logical interpretation derives new significance from empirical areas of cognitive neuroscience. By and large, the contribution of empirical findings is unmistakable in any sufficiently general theory of the logic of consciousness, quite independently of whether such a theory is intended for computational or philosophical purposes. To create realistic logics of knowledge and other propositional attitudes, one greatly benefits from a proper understanding of neural deficits, while the ensuing epistemic logics in turn clarify the actual neuropsychological content and processes of awareness, especially in relation to the implicit versus explicit distinction. For instance, one natural reason behind this distinction is the fact that explicit attitudes are not closed under logical consequence relation.

Acknowledgements

The work on this paper has been supported by the Osk. Huttunen Foundation, the Ella and Georg Ehrnrooth Foundation, and the Academy of Finland (project no: 17856, ‘Game-Theoretical Semantics and Its Applications’). I would like to thank two anonymous referees for commenting on the earlier version of this paper.

References


