Total Knowledge*

Ian Pratt-Hartmann
Department of Computer Science
University of Manchester,
Manchester M13 9PL, U.K.
ipratt@cs.man.ac.uk

Abstract

In this paper, we analyse a concept of total knowledge based on the idea that an agent's total knowledge is the strongest proposition the agent knows. We propose semantics for propositional and first-order languages with a modal operator TK representing total knowledge, and establish a result showing that total knowledge is 'epistemically categorical', in the sense that it determines the agent's knowledge over a broad range of contents. We show that (subject to some restrictions) total knowledge is always total knowledge of an objective content, and that, for such objective contents, our TK-operator corresponds in a straightforward way to Levesque's operator O.

Keywords: mathematical foundations, philosophical foundations, nonmonotonic reasoning.

Introduction

An agent that acquires information by the gradual accretion of propositions has finite knowldge: there is some proposition ϕ—the conjunction of all the propositions so-far acquired—which constitutes that agent's total knowledge. Since we can imagine situations in which it is useful for agents to reflect on their current epistemic states, it is natural to examine epistemic logics in which such states of total knowledge can be explicitly represented. That is the goal of the present paper.

To date, most research on representing total knowledge has focused on its role in reconstructing various forms of nonmonotonic logic. The origin of these ideas can be traced back to the original non-monotonic logic of (McDermott & Doyle 1980; 1982) and its later modifications e.g. in (Halpern & Moses 1985). However, the best-known such reconstruction is (Levesque 1990), extended and discussed in (Halpern & Lakemeyer 1995), (Lakemeyer 1993; 1996) and (Lakemeyer & Levesque 1998). For an overview of the relationships between these closely related approaches, see (Donini, Nardi, & Rosati 1997) and (Rosati 2000). Chen (1997) presents an analysis relating Levesque's concept of only knowing to the method of epistemic specifications of (Gelfond 1991).

*The author thanks Nick Player, Manfred Jaeger and Renate Schmidt for their comments on earlier drafts of this paper.

However, references to a proposition's being all that an agent knows also occur outside nonmonotonic logic, most notably, in discussions of probabilistic updating. For example, debates about the appropriateness of conditionalization as an updating strategy generally assume that probabilities are conditionalized on one's *total* knowledge: conditionalizing on just *part* of what one knows is (as far as the author is aware) never seriously proposed. But what does it mean, in this context, to say that a given proposition is one's total knowledge or total evidence? What are the implications of the assumption that such a proposition exists? Does this assumption affect the logic of knowledge in any way? Although there is much debate in the philosophical literature about the reasonableness of the assumption that evidence is propositional at all (see, e.g. (Jeffrey 1992), ch. 1), the implications for epistemic logic of the assumption that agents have (finite) total knowledge have been relatively neglected.

The goal of the present paper is to analyse a concept of total knowledge based on the intuition that an agent's total knowledge is the logically strongest proposition the agent knows, and to relate it to the corresponding concept employed by Levesque. In the course of our analysis, we will see that our concept of total knowledge shares many of the properties of Levesque's, though not the latter's central role in defeasible inference. This is a useful insight, because the concept presented here is arguably simpler and more intuitive than that used by Levesque, and may therefore be more appropriate in contexts other than the reconstruction of nonmonotonic inference. Certainly, the nontrivial and subtle nature of the relationship we map out illustrates the complexity and fecundity of the relevant concepts.

Total knowledge

The concept of total knowledge we will be working with is that of the strongest proposition an agent knows. Roughly, $TK\phi$ means that the agent knows ϕ, but does not know anything which knowing ϕ does not entail. This seems to be the most natural reconstruction of the concept of total knowledge appealed to when one is is enjoined to conditionalize on one's total knowledge.

Definition 1. Assume as given a countable set of *variables*, a countable set of *names* and, for each n ($0 \leq n$), a countable set of n-ary *predicate letters*. The symbol $=$ is one of the

binary predicate letters. We call the 0-ary predicate letters *proposition letters*. A *term* is a variable or a name.

Define the formulas of $\mathcal{FOLTK}$ to be the smallest set of expressions satisfying the following rules:

- if r is an n-ary predicate letter and $t_1, \ldots, t_n$ are terms, then $r(t_1, \ldots, t_n)$ is a formula of $\mathcal{FOLTK}$;
- if ϕ and ψ are formulas of $\mathcal{FOLTK}$ and x is a variable, then $\phi \wedge \psi$, $\phi \vee \psi$, $\neg\phi$, $\exists x\phi$, $\forall x\phi$ and $K\phi$ are formulas of $\mathcal{FOLTK}$;
- If ϕ is a formula of $\mathcal{FOLTK}$ and contains no occurrence of TK, then $TK\phi$ is a formula of $\mathcal{FOLTK}$.

Define the formulas of $\mathcal{PCKT}$ to be those formulas of $\mathcal{FOLTK}$ involving no occurrences of $\exists$ or $\forall$ and no n-ary relations for $n > 0$.

Formulas involving no occurrences of TK are called *basic*; formulas involving no occurrences of K or TK are called *objective*. Formulas in which every predicate letter appears within the scope of either K or TK are called *subjective*. The notion of a *free* occurrence of a variable is defined in the usual way. We use the connectives $\rightarrow$ and $\leftrightarrow$ as abbreviations with their usual meanings. A formula with no free variables is a *sentence*.

We have restricted the syntax of $\mathcal{FOLTK}$ so that TK may apply only to basic formulas. In fact, this restriction is inessential: all the theorems reported below hold even when it is lifted. However, we maintain it throughout most of this paper for the purpose of simplifying proofs. (We indicate inessential restrictions of theorems to basic formulas using parentheses.)

The general semantic framework used here is that of (Levesque 1990). Models for $\mathcal{FOLTK}$-formulas are sets of "interpretations", where an interpretation is just a model of the underlying nonmodal language. The most notable features are that names denote rigidly and uniquely, and that the domain of quantification is covered by the names. We have taken advantage of these features to simplify the statement of the semantics slightly, and we have made one additional, substantive change (discussed below).

Definition 2. An *interpretation* w is a function mapping any n-ary predicate letter r to a set r^w of n-tuples of names, subject to the constraint that $=^w$ is the identity relation on the set of names. (As usual, we assume that there is exactly one 0-tuple of names.)

Let W be a set of interpretations, let $w \in W$, and let ϕ be a sentence of $\mathcal{FOLTK}$. We define $W \models_w \phi$ inductively as follows:

- If r is a predicate letter and $a_1, \ldots, a_n$ are names, then $W \models_w r(a_1, \ldots, a_n)$ if and only if $a_1, \ldots, a_n \in r^w$;
- $W \models_w \phi \wedge \psi$ if and only if $W \models_w \phi$ and $W \models_w \psi$, and similarly for the other Boolean connectives;
- $W \models_w \exists x\phi$ if and only if $W \models_w \phi[x/a]$ for some name a, and similarly for the universal quantifier;
- $W \models_w K\phi$ if and only if, for all $w' \in W$, $W \models_{w'} \phi$;
- $\star$ $W \models_w TK\phi$ if $W \models_w K\phi$ and $W \models_w \neg K\chi$ for all objective sentences χ such that $\not\models K\phi \rightarrow \chi$.

Here, $\phi[x/a]$ denotes the result of substituting the name a for every free occurrence of x in ϕ.

Since only *sentences* receive truth-values, we will henceforth notate free variables explicitly. Thus, ϕ will denote a sentence, and $\psi(\bar{x})$ a formula with $\bar{x}$ as its only free variables. Note that if ϕ is objective, we can write $\models_w \phi$ for $W \models_w \phi$, and if ϕ is subjective, we can write $W \models \phi$ for $W \models_w \phi$. More generally, we write $W \models \phi$ to mean $W \models_w \phi$ for all $w \in W$, and $\models \phi$ to mean $W \models \phi$ for all W. We say ϕ is *consistent* if $W \models_w \phi$ for some W and some $w \in W$, and we say ϕ is *valid* if $\models \phi$. Clearly, the usual S5-axioms for K are valid.

At this point, we might pause to get a feel for our new operator by examining some of its salient properties. It is immediate from definition 2 that $\models TK\phi \rightarrow K\phi$ and, moreover, that

$$\text{if } \models K\phi \leftrightarrow K\psi \quad \text{then} \quad \models TK\phi \leftrightarrow TK\psi. \tag{1}$$

That is: if knowing ϕ and knowing ψ are the same state of affairs, then only knowing ϕ and only knowing ψ are also the same state of affairs. Moreover, since $\models K\phi \leftrightarrow KK\phi$, condition (1) has, as an immediate consequence

$$\models TK\phi \leftrightarrow TKK\psi. \tag{2}$$

Finally, anticipating a result proved below, it turns out that if we lift the restriction stating that TK may apply only to basic formulas, we obtain:

$$\models TK\phi \leftrightarrow TKTK\psi. \tag{3}$$

Properties (1)–(3) seem reasonable ones for a concept of total knowledge to exhibit, though, admittedly, intuition may be uncertain on the last of these. By contrast, if we consider Levesque's operator O (which corresponds roughly to our operator TK), we see that these properties fail. In particular, if ϕ is objective and is not logically true, then, on Levesque's semantics, $OK\phi$ and $OO\phi$ are both logically false. (At the same time, for any formula ϕ, $O(K\phi \wedge \phi)$ is logically equivalent to $O\phi$!) One of the surprising results of this paper is just how many features of Levesque's O-operator do nevertheless carry over to TK.

Let us return to the semantics of TK, given in clause $\star$ of definition 2. Observe that the quantification in this clause is restricted to *objective* sentences χ. (This restriction has nothing to do with our earlier syntactic stipulation that TK can apply only to basic formulas!) Allowing χ to range over *arbitrary* sentences in $\star$ would result in a nonterminating recursive definition of $\models$, since the truth of $TK\phi$ in W would depend on the truth of more complex sentences χ. Moreover, allowing χ to range over *basic* sentences in $\star$, though it would result in a well-formed definition, would have other undesirable consequences. Consider, for example, the sentence TKp_1. We do not want this sentence to be inconsistent, since it seems reasonable that an agent may have simply learned p_1 and nothing else. Yet Kp_1 fails to imply both p_2 and $\neg Kp_2$, so that, without the restriction of χ to objective sentences, clause $\star$ would make TKp_1 entail both $\neg Kp_2$ and $\neg K\neg Kp_2$, which is inconsistent on our semantics. Hence the restriction of χ to objective sentences in $\star$.

However, this restriction creates a problem. Consider the following consequence of $\star$.

Lemma 1. *For any (basic) sentence ϕ and any objective sentence χ, $\models TK\phi \rightarrow K\chi$ or $\models TK\phi \rightarrow \neg K\chi$.*

Proof. If $\models K\phi \rightarrow \chi$ then $\models TK\phi \rightarrow K\chi$. □

Lemma 1 states that, as we might say, $TK\phi$ is *epistemically categorical* for objective sentences χ. Yet we would prefer that $TK\phi$ were epistemically categorical for *arbitrary* χ. After all, an agent's total knowledge should determine exactly what the agent does and does not know. One of the main results about the TK operator is that, in the current semantic framework, lemma 1 can be strengthened in just this way. Again, to simplify the proofs, we restrict the result in this paper to basic χ.

It is worth pausing to see why this result is surprising. Lemma 1 guarantees that any two agents whose total knowledge is ϕ know the same objective sentences. However, it is easy to construct an example of two agents who know the same objective sentences but who do not know the same basic sentences. Let p be a unary predicate letter, and enumerate the names as $\{c_i\}_{0\leq i}$. Define the interpretation w_0 by setting $\models_{w_0} p(c_j)$ if and only if j is odd; and define the interpretation w_i, for $i \geq 1$ by setting $\models_{w_i} p(c_j)$ if and only if j is odd or $j = 2i$. Assume that all other predicate letters are assigned the empty interpretation. Let $W = \{w_i | i \geq 0\}$ and $W' = \{w_i | i \geq 1\}$. Then it is easy to see that, for all objective χ, $W \models K\chi$ if and only if $W' \models K\chi$. (For a sketch proof, see (Levesque 1990), lemma 3.6.2.) However, we have $W' \models K\exists x(p(x) \wedge \neg Kp(x))$ but $W \models \neg K\exists x(p(x) \wedge \neg Kp(x))$. The analysis below shows that this sort of situation cannot arise in the presence of total knowledge.

The propositional case

We begin with a simple observation establishing the consistency of certain total-knowledge sentences.

Lemma 2. *If a sentence ϕ of $\mathcal{FOLTK}$ is objective and consistent, then $TK\phi$ is consistent.*

Proof. For each objective χ such that $\not\models K\phi \rightarrow \chi$, we have $\not\models \phi \rightarrow \chi$, so let w_χ be an interpretation such that $\models_{w_\chi} \phi \wedge \neg\chi$. Let W be the set consisting of all these w_χ. Since ϕ is consistent, $W \neq \emptyset$, and it is easy to see that $W \models TK\phi$. □

The analysis of TK in the propositional case is very easy, and relies on the existence of the following normal-form theorem.

Lemma 3. *Any basic sentence of $\mathcal{PCTK}$ is equivalent to a sentence of the form*

$$\bigvee_{1\leq h\leq l} (K\psi_h \wedge \neg K\chi_{h,1} \wedge \ldots \wedge \neg K\chi_{h,m_h} \wedge \pi_h).$$

in which the ψ_h, $\chi_{h,i}$ and π_h are objective.

Proof. Straightforward from standard S5-identities. □

Thus, K-operators occurring in the scope of other K-operators in basic $\mathcal{PCTK}$ sentences can always be eliminated. Of course, $\mathcal{FOLTK}$ lacks this feature: the embedded K in $K\exists x(p(x) \wedge \neg Kp(x))$ cannot be removed.

As a corollary of this normal form lemma, we have

Lemma 4. *Let ϕ be a consistent (basic) sentence of $\mathcal{PCTK}$. Then there exists a basic (in fact, objective) sentence ψ, such that $\phi \wedge TK\psi$ is consistent.*

Proof. Assume without loss of generality that ϕ is of the form given in lemma 3, with the first disjunct consistent. Then $\not\models K\psi_1 \rightarrow \chi_{1,j}$ for all j ($1 \leq j \leq m_1$), and $\not\models \psi_1 \rightarrow \neg\pi_1$.

Now consider $TK\psi_1$. This sentence is consistent by lemma 2. Moreover, $\not\models K\psi_1 \rightarrow \chi_{1,j}$ implies $\models TK\psi_1 \rightarrow \neg K\chi_{1,j}$. Finally, since the objective sentence $\psi_1 \wedge \pi_1$ is true in some interpretation w, if $W \models TK\psi_1$, then it is easy to see that $W \cup \{w\} \models_w \pi_1 \wedge TK\psi_1$. Hence $\phi \wedge TK\psi_1$ is consistent. □

Lemma 4 ensures that, in the propositional case, the assumption that there is a sentence which is the agent's total knowledge does not change the finitary logic of knowledge: any (basic) sentence which is consistent without this assumption is consistent in its presence. However, we show below that lemma 4 is false for $\mathcal{FOLTK}$.

The first-order case

The following construction is crucial in understanding the behaviour of TK in the first-order case.

Definition 3. A *permutation of individuals* is a function from the set of names to the set of names which is 1–1 and onto. If f is a permutation of individuals, then it is extended to apply to interpretations and formulas as follows. If w is an interpretation, for any n-ary predicate letter r, let $a_1, \ldots a_n \in r^{f(w)}$ if and only if $f^{-1}(a_1), \ldots, f^{-1}(a_n) \in r^w$. If x is a variable, let $f(x) = x$. If $r(t_1, \ldots t_n)$ is an atomic formula, let $f(r(t_1, \ldots t_n)) = r(f(t_1), \ldots f(t_n))$, and let f be defined on nonatomic formulas by $f(\phi \wedge \psi) = f(\phi) \wedge f(\psi)$, $f(\phi \vee \psi) = f(\phi) \vee f(\psi)$, $f(\neg\phi) = \neg f(\phi)$, $f(\exists x\phi) = \exists x f(\phi)$, $f(\forall x\phi) = \forall x f(\phi)$, $f(K\phi) = Kf(\phi)$, $f(TK\phi) = TKf(\phi)$.

Thus, when applying f to interpretations and formulas, we switch round the extensions of predicates and the names occurring in formulas in corresponding ways.

Lemma 5. *If f is a permutation of individuals, then f is also 1–1 and onto on the set of interpretations, the set of formulas, the set of basic formulas and the set of objective formulas. Furthermore, for all sentences ϕ, sets of interpretations W and interpretations $w \in W$, $W \models_w \phi$ if and only if $f(W) \models_{f(w)} f(\phi)$.*

Proof. The first part of the lemma is obvious. The second part follows by structural induction on ϕ. □

Definition 4. Let $\bar{x} = x_1, \ldots x_n$ be a tuple of variables with X the set $\{x_1, \ldots, x_n\}$. Let $A = \{a_1, \ldots, a_m\}$ (with the a_i distinct) be a set of names. Let $P_1, \ldots, P_m$ be a set of (possibly empty) disjoint subsets of X and let $P_{m+1}, \ldots, P_{m+l}$

be a partition of $X \setminus \bigcup_{1 \leq i \leq m} P_i$. (Thus, $0 \leq l \leq n$.) A *distribution formula* (for $\bar{x}$ and A) is a consistent formula of the form $\delta(\bar{x}) :=$

$\bigwedge\{x_j = a_i | 1 \leq i \leq m, 1 \leq j \leq n, \text{ and } x_j \in P_i\}$
$\bigwedge\{x_j = x_k | m+1 \leq i \leq m+l, \text{ and } x_j, x_k \in P_i\}$
$\bigwedge\{x_j \neq a_i | 1 \leq i \leq m, m+1 \leq i' \leq m+l, \text{ and } x_j \in P_{i'}\}$
$\bigwedge\{x_j \neq x_k | m+1 \leq i < i' \leq m+l,$
$x_j \in P_i \text{ and } x_k \in P_{i'}\}.$

For a given $\bar{x}$ and A, denote the set of all such formulas by $\Delta_A(\bar{x})$. If $n = 0$, set $\Delta_A = \{\top\}$.

Intuitively, $\delta(\bar{x})$ assigns every variable in $\bar{x}$ to one of $m+l$ 'boxes'. Variables assigned to the same box are asserted to be identical and variables assigned to different boxes are asserted to be distinct. Variables assigned to box i ($1 \leq i \leq m$) are asserted to be identical to a_i.

Lemma 6. *Let $\bar{x} = x_1, \ldots x_n$ be a tuple of variables and A a set of names. Then $\Delta_A(\bar{x})$ is a partition. That is: $\models \forall \bar{x} \bigvee \Delta_A(\bar{x})$, and $\models \forall \bar{x} \neg(\delta(\bar{x}) \wedge \delta'(\bar{x}))$ for distinct $\delta(\bar{x}), \delta'(\bar{x}) \in \Delta_A(\bar{x})$.*

Proof. Obvious. □

We note that distribution formulas are *rigid*: they are satisfied by the same tuples regardless of the interpretation. Hence we sometimes write $\models \delta(\bar{a})$ without mentioning W or w.

Lemma 7. *Let ϕ be a sentence and $\psi(\bar{x})$ a formula. Let C be the set of names occurring in either formula. Then there exists a disjunction $\pi(\bar{x})$ of formulas in $\Delta_C(\bar{x})$ such that, for all tuples $\bar{a}$, $\models \pi(\bar{a})$ if and only if $\models \phi \rightarrow \psi(\bar{a})$.*

Proof. Suppose that $\bar{a}$ and $\bar{a}'$ satisfy the same $\delta(\bar{x})$ in $\Delta_C(\bar{x})$. Then the mapping $\bar{a} \mapsto \bar{a}'$ is well-defined and extends to a permutation of individuals f such that f is the identity on C. Hence $f(\phi) = \phi$ and $f(\psi(\bar{a})) = \psi(\bar{a}')$. By lemma 5, $\models \phi \rightarrow \psi(\bar{a})$ if and only if $\models \phi \rightarrow \psi(\bar{a}')$. Now set $\pi(\bar{x}) :=$

$$\bigvee\{\delta(\bar{x}) \in \Delta_C(\bar{x}) : \models \delta(\bar{a}') \text{ for some } \bar{a}' \text{ s.t. } \models \phi \rightarrow \psi(\bar{a}')\}.$$

(As usual, we take $\bigvee \emptyset$ to be $\bot$.) Suppose $\models \phi \rightarrow \psi(\bar{a})$. Since $\Delta_C(\bar{x})$ is a partition, $\models \delta(\bar{a})$ for some $\delta(\bar{x})$, so $\models \pi(\bar{a})$. Conversely, suppose $\models \pi(\bar{a})$. Then $\models \delta(\bar{a})$ for some $\delta(\bar{x})$, such that, for some $\bar{a}'$, $\models \delta(\bar{a}')$ and $\models \phi \rightarrow \psi(\bar{a}')$. But since $\bar{a}$ and $\bar{a}'$ satisfy the same $\delta(\bar{x})$ in $\Delta_C(\bar{x})$, we have $\models \phi \rightarrow \psi(\bar{a})$. □

Lemma 8. *Let ϕ be a (basic) sentence and $\psi(\bar{x})$ an objective formula. Let C be the set of names occurring in either formula. Then there exists a disjunction $\pi(\bar{x})$ of formulas in $\Delta_C(\bar{x})$ such that $\models TK\phi \rightarrow \forall \bar{x}(K\psi(\bar{x}) \leftrightarrow \pi(\bar{x}))$.*

Proof. By lemma 7, let $\pi(\bar{x})$ be such that, for all tuples $\bar{a}$, $\models \pi(\bar{a})$ if and only if $\models K\phi \rightarrow \psi(\bar{a})$. Let W be any set of interpretations and let $\bar{a}$ be any tuple. If $W \models TK\phi$, then, by the semantics of TK, $W \models K\psi(\bar{a})$ if and only if $\models K\phi \rightarrow \psi(\bar{a})$. The result is then immediate. □

Theorem 1. *Let ϕ be a (basic) sentence and $\psi(\bar{x})$ a (basic) formula. Then there is a disjunction $\pi(\bar{x})$ of elements of $\Delta_C(\bar{x})$ for some C, such that $\models TK\phi \rightarrow \forall \bar{x}(K\psi(\bar{x}) \leftrightarrow \pi(\bar{x}))$.*

Proof. We proceed by induction on the number n of occurrences of K in $\psi(\bar{x})$. The case $n = 0$ is handled by lemma 8. If $n > 0$, let $K\psi'(\bar{x}')$ be a subformula of $\psi(\bar{x})$, with $\psi'(\bar{x}')$ objective. By lemma 8, let $\pi'(\bar{x}')$ be such that $\models TK\phi \rightarrow \forall \bar{x}'(K\psi'(\bar{x}') \leftrightarrow \pi'(\bar{x}'))$. and let $\psi''(\bar{x})$ be the result of substituting $\pi'(\bar{x}')$ for $K\psi'(\bar{x}')$ in $\psi(\bar{x})$. Then, $\models TK\phi \rightarrow \forall \bar{x}(K\psi(\bar{x}) \leftrightarrow K\psi''(\bar{x}))$. Since $\psi''(\bar{x})$ has fewer than n occurrences of K, the result follows by inductive hypothesis. □

Note that this straightforward induction depends on the fact that $\psi(\bar{x})$ is basic. This is because, in any set of interpretations W, the truth-values of $K\psi(\bar{a})$ and $K\psi''(\bar{a})$ depend only on the truth-values of their subformulas *at the worlds in* W. Since $\pi'(\bar{x}')$ and $K\psi'(\bar{x}')$ are satisfied by the same tuples in any world of W, it is obvious that $\psi(\bar{a})$ and $\psi''(\bar{a})$ must have the same truth value in every world of W as well. However, such a substitution within the scope of TK-operators would in general not be truth-preserving. (As stated above, theorem 1 does in fact hold for arbitrary $\psi(\bar{x})$; however, the proof in this case is more delicate.)

Corollary 1. *For all (basic) sentences ϕ and ψ, $\models TK\phi \rightarrow K\psi$ or $\models TK\phi \rightarrow \neg K\psi$.*

Proof. By theorem 1, $\models TK\phi \rightarrow \forall \bar{x}(K\psi \leftrightarrow \pi)$, where π is a disjunction of elements of Δ_C for some C (with a 0-tuple of variables). Hence π is $\bot$ or $\top$. □

Corollary 2. *Let ϕ be a (basic) sentence and $\psi(x)$ a basic formula with one free variable. Suppose that $W \models TK\phi$. Then the set $\{a : W \models K\psi(a)\}$ is finite or cofinite.*

Proof. By theorem 1, $\models TK\phi \rightarrow \forall x(K\psi(x) \leftrightarrow \pi(x))$, where $\pi(x)$ is a disjunction of elements of $\Delta_C(x)$ for some C (with a single variable x). Clearly, the set of a satisfying $\pi(x)$ is finite or cofinite. □

Recall that, in the propositional case, if ϕ is consistent, then we can find ψ such that $\phi \wedge TK\psi$ is consistent. In the first-order case, this is no longer true.

Theorem 2. *There exists a consistent basic sentence ϕ such that, for all (basic) sentences ψ, $\models \phi \rightarrow \neg TK\psi$.*

Proof. If $\psi'(x)$ is any formula with one free variable x, let $\exists_\infty x\psi'(x)$ abbreviate some sentence or other implying that $\psi'(x)$ is satisfied by infinitely many values of x. Let $p(x)$ be a unary predicate letter, and let ϕ be a consistent basic sentence of the form $\exists_\infty x Kp(x) \wedge \exists_\infty x \neg Kp(x)$. It is easy to see that such a ϕ can be found. By corollary 2, $\models \phi \rightarrow \neg TK\psi$ for all basic sentences ψ. □

Thus, in the first-order case, the assumption that there is total knowledge changes the finitary logic of knowledge: basic sentences that are consistent without this assumption may be inconsistent in its presence.

Next, we show that total knowledge of any (basic) sentence is logically equivalent to total knowledge of an objective sentence. We need the following general lemma.

Lemma 9. *Let ϕ and ψ be (basic) sentences such that $\models TK\phi \rightarrow K\psi$, $\models K\psi \rightarrow K\phi$ and $TK\phi$ is consistent. Then $\models TK\phi \leftrightarrow TK\psi$.*

Proof. Suppose $W \models TK\phi$. Then $W \models K\psi$. Let χ be objective with $\not\models K\psi \rightarrow \chi$. Then $\not\models K\phi \rightarrow \chi$, because $\models K\psi \rightarrow K\phi$. So $W \models \neg K\chi$. Hence $W \models TK\psi$.

Conversely, suppose $W \models TK\psi$. Then $W \models K\phi$. Let χ be objective with $\not\models K\phi \rightarrow \chi$, so that $\models TK\phi \rightarrow \neg K\chi$. Then $\not\models K\psi \rightarrow \chi$ also, since otherwise, given that $\models TK\phi \rightarrow K\psi$, we would have $\models TK\phi \rightarrow K\chi$, contradicting the hypothesised consistency of $TK\phi$. But if $\not\models K\psi \rightarrow \chi$, then $W \models \neg K\chi$. Hence $W \models TK\phi$. □

Theorem 3. *Let ϕ be a (basic) sentence. Then there exists an objective sentence ϕ^* such that $\models TK\phi \leftrightarrow TK\phi^*$.*

Proof. If ϕ is already objective or if $TK\phi$ is inconsistent, the result is trivial, so we may assume otherwise. Let $K\psi_1(\bar{x}_1)$ be a subformula of ϕ, with $\psi_1(\bar{x}_1)$ objective. Then we can find ρ_1 such that $\models \phi \rightarrow \rho_1$, where ρ_1 is the sentence $\forall\bar{x}_1(K\psi(\bar{x}_1) \leftrightarrow \pi_1(\bar{x}))$ constructed as in lemma 8. Let ϕ_1 be the result of substituting $\pi_1(\bar{x})$ for $K\psi(\bar{x}_1)$ in ϕ. By lemma 8, $\models TK\phi \rightarrow K(\phi_1 \wedge \rho_1)$, and certainly $\models K(\phi_1 \wedge \rho_1) \rightarrow K\phi$. Since $TK\phi$ is assumed consistent, lemma 9 implies that $\models TK\phi \leftrightarrow TK(\phi_1 \wedge \rho_1)$. If there is a subformula $K\psi_2(\bar{x}_2)$ in ϕ_1 with $\psi_2(\bar{x}_2)$ objective, we proceed as before, obtaining $\models TK\phi \leftrightarrow TK(\phi_2 \wedge \rho_1 \wedge \rho_2)$, and so on, until we eventually obtain $\models TK\phi \leftrightarrow TK(\phi_m \wedge \rho_1 \wedge \ldots \wedge \rho_m)$, with ϕ_m objective and $m \geq 1$.

Now consider in more detail the sentence $\rho_1 \wedge \ldots \wedge \rho_m$. Ignoring the previous numbering, this may be written out as a conjunction of the form $\bigwedge_{1\leq j\leq M} \forall\bar{x}_j(\delta_j(\bar{x}_j) \rightarrow K\psi_j(\bar{x}_j)) \wedge \bigwedge_{1\leq j\leq M'} \forall\bar{x}'_j(\delta'_j(\bar{x}'_j) \rightarrow \neg K\psi'_j(\bar{x}'_j))$ where the $\delta_j(\bar{x}_j)$, $\delta'_j(\bar{x}'_j)$ are conjunctions of equality and inequality formulas, and the $\psi_j(\bar{x}_j)$, $\psi'_j(\bar{x}'_j)$ are objective. Since the $\delta_j(\bar{x}_j)$ are in fact rigid, we have

$$\models K\forall\bar{x}_j(\delta_j(\bar{x}_j) \rightarrow K\psi_j(\bar{x}_j)) \leftrightarrow K\forall\bar{x}_j(\delta_j(\bar{x}_j) \rightarrow \psi_j(\bar{x}_j)).$$

Hence we can omit the K from the relevant conjuncts and set ϕ^* to be

$$\phi_m \wedge \bigwedge_{1\leq j\leq M} \forall\bar{x}_j(\delta_j(\bar{x}_j) \rightarrow \psi_j(\bar{x}_j)),$$

whence $\models TK\phi \leftrightarrow TK(\phi^* \wedge \sigma_1 \wedge \ldots \wedge \sigma_{M'})$ where σ_j is $\forall\bar{x}'_j(\delta'_j(\bar{x}'_j) \rightarrow \neg K\psi'_j(\bar{x}'_j))$.

To complete the proof, suppose $\bar{a}$ is a tuple with $\models \delta'_j(\bar{a})$. Since $TK\phi$ is consistent, $\not\models K\phi^* \rightarrow \psi'_j(\bar{a})$. Hence, since $\psi'_j(\bar{a})$ is objective, $\models TK\phi^* \rightarrow \neg K\psi'_j(\bar{a})$. Thus, $\models TK\phi^* \rightarrow \forall\bar{x}'_j(\delta'_j(\bar{x}'_j) \rightarrow \neg K\psi'_j(\bar{x}'_j))$. Hence, we have $\models TK\phi^* \rightarrow K(\phi^* \wedge \sigma_1 \wedge \ldots \wedge \sigma_{M'})$, $\models K(\phi^* \wedge \sigma_1 \wedge \ldots \wedge \sigma_{M'}) \rightarrow K\phi^*$ and finally, by lemma 2, $TK\phi^*$ consistent. By lemma 9, $\models TK\phi^* \leftrightarrow TK(\phi^* \wedge \sigma_1 \wedge \ldots \wedge \sigma_{M'})$, and we are done. □

Comparison with only knowing

An alternative approach to total knowledge is provided by (Levesque 1990). Before we give the semantics for Levesque's operator, we need to mention a difference between Levesque's basic formalism and the one adopted in this paper. So far, we have assumed that, in an assertion of the form $W \models_w \phi$, w is a member of W. But in fact, the definitions work perfectly well without this assumption, the major effect being that $K\phi \wedge \neg\phi$ becomes satisfiable. (Levesque actually uses the letter B where we have used K.) Given this change, Levesque can give the semantics of the modal operator O as:

$W \models_w O\phi$ if and only if $W \models K\phi$ and, for all w such that $W \models_w \phi$, $w \in W$.

The semantics for K and the nonmodal connectives are unaffected.

Levesque's semantics for O have the desired effect only when the set of interpretations W is maximal in the following sense:

Definition 5. Let W and W' be sets of interpretations. We say that W and W' are *equivalent* if, for all basic sentences ϕ, $W \models K\phi$ if and only if $W' \models K\phi$.

A set of interpretations W is *maximal* if, for all W' such that $W \equiv W'$ and $W \subseteq W'$, we have $W = W'$.

The motivation for this definition is that, if W is a set of interpretations and $w \in W$ is an interpretation such that $W \models_w \phi$, then it can turn out that $O\phi$ is true in W and false in $W \setminus \{w\}$, even though and W and $W \setminus \{w\}$ give the agent the same basic beliefs! By ignoring nonmaximal sets W, this anomaly is avoided.

Theorem 4. *Let ϕ be objective and let $W \neq \emptyset$ be any set of interpretations (not necessarily maximal) such that $W \models O\phi$. Then $W \models TK\phi$. Conversely, Let ϕ be objective and let $W \neq \emptyset$ be a maximal set of interpretations such that $W \models TK\phi$. Then $W \models O\phi$.*

Proof. For the first part, we certainly have $W \models K\phi$. Moreover, let χ be objective with $\not\models K\phi \rightarrow \chi$. Certainly, then $\not\models \phi \rightarrow \chi$. So let w be an interpretation such that $\models_w \phi \wedge \neg\chi$. Since $W \models O\phi$, we have $w \in W$, whence $W \models \neg K\chi$. Thus, $W \models TK\phi$.

For the second part, again we certainly have $W \models K\phi$. Moreover, let w be an interpretation such that $\models_w \phi$. Suppose ψ is any basic sentence such that $W \models K\psi$. Since $W \models TK\phi$, it follows from corollary 1 that $\models TK\phi \rightarrow K\psi$. Now ϕ is objective, $\models_w \phi$ and $W \models TK\phi$, so $W \cup \{w\} \models TK\phi$, and so $W \cup \{w\} \models K\psi$. Thus, for any basic ψ, $W \models K\psi$ implies $W \cup \{w\} \models K\psi$. This easily implies that, for any basic ψ, $W \models K\psi$ if and only if $W \cup \{w\} \models K\psi$. That is, $W \equiv W \cup \{w\}$. By the maximality of W, then, $w \in W$, and hence $W \models O\phi$. □

Note that the second part of the above theorem depends crucially on the strengthening of lemma 1 provided by corollary 1.

It is easy to construct examples showing that theorem 4 fails if ϕ is allowed to be nonobjective. Consider for example the sentence $\phi := \neg Kp \rightarrow q$. The sentence $O\phi$ is

consistent, and implies Kq. (Thus, ϕ can be seen as a default rule licencing inference to q provided p is not known.) By contrast, $TK\phi$ is easily seen to be inconsistent, since $K\phi$ implies neither p nor q, so that $TK\phi$ implies the inconsistent trio $\neg Kp, \neg Kq, K(\neg Kp \rightarrow q)$.

This last example shows how the failure of property (1) above is crucial for default inference. By simple S5-manipulation,

$$\models K(\neg Kp \rightarrow q) \leftrightarrow K(Kp \vee Kq)$$

and so by property (1),

$$\models TK(\neg Kp \rightarrow q) \leftrightarrow TK(Kp \vee Kq).$$

But the formula $(Kp \vee Kq)$ is symmetric in p and q, and thus could not possibly favour inferring Kq over inferring Kp. Thus, no concept of total knowledge for which property (1) obtains is likely to be of any use for modelling default inference along the lines taken by Levesque.

As we have already remarked, the restriction in the final clause of definition 1 that TK applies only to formulas not involving any occurrences of TK is inessential. The semantics presented in definition 2 work unproblematically even when it is lifted.

The following result is immediate from the semantics for TK and K.

Lemma 10. *If ϕ is any sentence, then $\models TK\phi \rightarrow KTK\phi$.*

We note that the proof of lemma 9 does not depend on any assumption that ϕ and ψ are basic, so that the result holds for all ϕ and ψ. We then have

Corollary 3. *For any formula ϕ, $\models TK\phi \leftrightarrow TKTK\phi$.*

Proof. If $TK\phi$ is inconsistent, then $TKTK\phi$ is certainly inconsistent. Hence we may assume that $TK\phi$ is consistent. We have $\models TK\phi \rightarrow KTK\phi$ by lemma 10, and certainly $\models KTK\phi \rightarrow K\phi$. Hence, by lemma 9, putting $\psi := TK\phi$, we have $\models TK\phi \leftrightarrow TKTK\phi$. □

This is the promised proof of the property (3) above.

Conclusions and further work

The purpose of this paper has been to define and analyse a concept of total knowledge based on the idea that an agent's total knowledge is the strongest proposition that the agent knows. We proposed semantics for the languages $\mathcal{PCTK}$ and $\mathcal{FOLTK}$, according to which a sentence $TK\phi$ was guaranteed to be epistemically categorical for objective sentences. We showed that, surprisingly, total knowledge is epistemically categorical for all basic sentences. We showed that the assumption that an agent has total knowledge does not change the finitary logic of $\mathcal{PCTK}$; but it does change the finitary logic of $\mathcal{FOLTK}$. We showed that total knowledge of any basic sentence is logically equivalent to total knowledge of some objective sentence. Finally, we showed that, for objective sentences, but not for nonobjective sentences, TK coincides with Levesque's operator O, modulo certain technical details.

The above results can be extended in several ways. Throughout most of this paper, we have assumed that TK-operators could apply only to basic formulas. In fact, this assumption is unnecessary, and all of the above theorems remain true when it is removed. The proofs cannot be presented within the confines of this paper. Another important extension is to index the modal operators K and TK to indicate the time at which the knowledge (or total knowledge) applies. Thus, we might work instead with operators K_n ("I know at time n that ... ...") and TK_n ("My total knowledge at time n is that ... ..."). The extension of the semantics to these temporally indexed cases is routine. It turns out that lemma 4 continues to hold for the temporally indexed version of $\mathcal{PCTK}$. The proof is more involved than in the nontemporal case, and cannot be given here.

References

Chen, J. 1997. The generalized logic of only knowing (GOL) that covers the notion of epistemic specifications. *Journal of Logic and Computation* 7(2):159–174.

Donini, F. M.; Nardi, D.; and Rosati, R. 1997. Ground nonmonotonic modal logics. *Journal of Logic and Computation* 7(4):523–548.

Gelfond, M. 1991. Strong introspection. In *Proceedings, AAAI*, 386–391. Los Altos, CA: Morgan Kaufmann.

Halpern, J. Y., and Lakemeyer, G. 1995. Levesque's axiomatization of only knowing is incomplete. *Artificial Intelligence* 74(2):381–387.

Halpern, J., and Moses, Y. 1985. Towards a theory of knowledge and ignorance: Preliminary report. In Apt, K., ed., *Logic and Models of Concurrent Systems*. Berlin: Springer. 459–476.

Jeffrey, R. C. 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Lakemeyer, G., and Levesque, H. J. 1998. AOL: a logic of acting, sensing, knowing and only knowing. In *Proceedings of the sixth International Conference on Principles of Knowledge Representation and Reasoning*.

Lakemeyer, G. 1993. All they know: a study in multi-agent autoepistemic reasoning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, 376–381.

Lakemeyer, G. 1996. Only knowing in the situation calculus. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning*, 14–25.

Levesque, H. J. 1990. All I know: a study in autoepistemic logic. *Artificial Intelligence*.

McDermott, D., and Doyle, J. 1980. Non-monotonic logic I. *Artificial Intelligence Journal* 13:41–72.

McDermott, D., and Doyle, J. 1982. Non-monotonic logic II: Non-monotonic modal theories. *Journal of the ACM* 29:33–57.

Rosati, R. 2000. On the decidability and complexity of reasoning about only knowing. *Artificial Intelligence* 116(1–2):193–215.