All You Ever Wanted to Know about Tweety*

Gerhard Lakemeyer
Institut für Informatik III
Universität Bonn
gerhard@uran.informatik.uni-bonn.de

Abstract

We propose a formalization of only-knowing-about, which captures the idea that something is all an agent knows about some subject matter. The work extends a previous formalization of only-knowing by Levesque. Besides discussing some of the logical properties of the new concept, we also address the issue of computing what is known about some subject matter for a given knowledge base. In this context, we are able to relate only-knowing-about to deKleer's ATMS. Finally, we show that only-knowing-about is efficiently computable, if we weaken the underlying model of belief.

1 All I know about...

In [Lev90], Levesque proposes a formalization of the concept of *only-knowing*. This notion, which captures the idea that something is *all* an agent knows,[1] has proven very valuable in characterizing the knowledge of a fully introspective agent. In particular, Levesque uses it to reconstruct and in fact generalize Moore's autoepistemic logic within the framework of a classical (monotonic) logic.

While a logic of only-knowing is thus very useful from a theoretical point of view, being able to talk about *all* an agent knows does not seem very relevant from a practical point of view, say, as part of a query language. After all, hardly anybody is interested in all I know.[2] A much more useful concept in this regard seems to be the notion of *only-knowing-about*, where one is interested in all the agent knows about a certain subject matter. Thus queries like *is this all you know about Tweety* or *what do you know about Tweety* seem very appropriate.

Another application is a multi-agent scenario where agents reason about other agents, possibly applying defaults. For example, assume there are two agents, Jack and Jill, with the default that birds fly unless known otherwise being common knowledge. If Jill knows

*This paper also appeared in: *Proc. of the 3rd Int. Conf. on Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, 1992, pp. 639–648.

[1]We will freely use the terms *knowledge* and *belief* interchangeably. While all of the logics discussed here allow an agent to have false beliefs, none of our results hinges on this choice.

[2]Besides, for a knowledge-based agent it is a trivial matter to tell a user all it knows. It only needs to hand over a copy of its knowledge base.

that all Jack knows about Tweety is that he is a bird, then Jill knows that Jack knows that Tweety flies. Notice that, while it is unreasonable to assume that an agent knows all another agent knows, it is quite possible that an agent knows all another agent knows about a certain subject. For example, Jill may just have bought Tweety and now tells Jack that Tweety is a bird.

In this paper, we propose a formalization of *only-knowing-about* in the propositional case. We do so by extending the existing logic of *only-knowing* developed by Levesque [Lev90]. Besides discussing general properties of only-knowing-about, we also address the issue of computing all an agent knows about a subject matter and relate it to deKleer's ATMS [deK86]. Finally, we show how, by weakening the underlying model of belief, this computation can be carried out efficiently.

To our knowledge, this is the first formalization of only-knowing-about that has been studied in detail. Levesque, in his paper on only-knowing, also suggests a definition of only-knowing-about which unfortunately has serious deficiencies. We will get back to this later. Preliminary results of our work appeared in [Lak92].

The paper is organized as follows. In Section 2, we introduce Levesque's logic of only-knowing. In Section 3, we extend Levesque's logic by introducing the concept of only-knowing-about and discuss some of its properties. Section 4 addresses the issue of computing all that is known about some subject matter for a given knowledge base. Section 5 shows how all that is known about some subject matter can be computed efficiently if we weaken the underlying model of belief appropriately. Finally, Section 6 summarizes the results and discusses future directions.

2 The Logic *OL*

In this section, we introduce a propositional version of Levesque's logic of only-knowing.[3]

The primitives of the language are a countably infinite set $\mathcal{P}$ of atomic propositions (or atoms), the connectives $\vee$, $\neg$, and the modal operators $\mathbf{L}$ and $\mathbf{O}$. For convenience we also include a special atom $\Box$, which always denotes falsity. Sentences are formed in the usual way from these primitives.[4]

Literals are either atoms or negated atoms. **Clauses** are disjunctions of literals. A clause c is contained in a clause c' ($c \subseteq c'$) if every literal in c other than $\Box$ occurs in c'. We write $c \subsetneq c'$ instead of $c \subseteq c'$ and $c' \not\subseteq c$. A sentence is called **objective** if it contains no modal operator, **subjective** if every atom occurs within the scope of a modal operator, and **basic** if it contains no $\mathbf{O}$'s.

The semantics of *OL* is based on the notion of a world, which is simply a truth assignment of the atoms.

Definition 1 (Worlds) *A world w is a function $w : \mathcal{P} \longrightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $w(\Box) = \mathbf{f}$.*

Belief ($\mathbf{L}$) is interpreted possible-world style.[5] The idea is that an agent imagines a set of

[3] Levesque considers the more general case of a first-order language with quantifying-in.

[4] We will freely use other connectives like $\wedge$, $\supset$ and $\equiv$, which should be understood as syntactic abbreviations of the usual kind.

[5] Possible-world semantics goes back to Kripke [Kri63]. Its application to epistemic notions is due to Hintikka [Hin62].

worlds all of which are compatible with what the agent believes about the world. In other words, the agent believes a sentence just in case that sentence is true in all the worlds he or she imagines.

In order for $\mathbf{O}\alpha$ to capture the intuition that α is all that is believed, we require that α not only holds at all the worlds that are imagined, i.e., that α is believed, but also that the set of worlds is as large as possible. In other words, we could not possibly add more worlds to the set and still believe α. If α is objective, there is a unique M that has this property, namely the set of *all* worlds where α is true. If α itself contains modal operators, there may be multiple M or none at all, a phenomenon that corresponds precisely to the fact that a sentence may have multiple stable expansions or none at all [Lev90].

Let p be an atom and let α and β be arbitrary sentences, The truth of a sentence α with respect to a world w and a set of worlds M ($M, w \models \alpha$) is defined as follows:

$$
\begin{array}{lcl}
M, w \models p & \iff & w(p) = \mathbf{t} \\
M, w \models \neg\alpha & \iff & M, w \not\models \alpha \\
M, w \models \alpha \vee \beta & \iff & M, w \models \alpha \text{ or } M, w \models \beta \\
M, w \models \mathbf{L}\alpha & \iff & \text{for all } w', \text{ if } w' \in M \text{ then } M, w' \models \alpha \\
M, w \models \mathbf{O}\alpha & \iff & \text{for all } w', \ w' \in M \text{ iff } M, w' \models \alpha
\end{array}
$$

Note that the only difference between $\mathbf{L}$ and $\mathbf{O}$ is that the *if ... then* is replaced by *iff.* As a notational convenience, we often write $M \models \alpha$ if α is subjective, since the truth of α depends solely on the set of worlds M. Similarly, we write $w \models \alpha$ for objective α.

A set of sentences Γ *logically implies* a sentence α ($\Gamma \models \alpha$) iff for all worlds w and sets of worlds M, if $M, w \models \gamma$ for all $\gamma \in \Gamma$, then $M, w \models \alpha$. α is *valid* ($\models \alpha$) iff $\{\} \models \alpha$.

2.1 Some properties of *OL*

The reader is referred to [Lev90] for a sound and complete axiomatization of *OL*. Here we only touch on some of the main aspects of the logic.

It is well known that, given a globally accessible set of worlds M, the properties of $\mathbf{L}$ are precisely those of *weak S5* [HM85], that is, beliefs are closed under logical implication ($\models \mathbf{L}\alpha \wedge \mathbf{L}(\alpha \supset \beta) \supset \mathbf{L}\beta$) and positive as well as negative introspection ($\models \mathbf{L}\alpha \supset \mathbf{LL}\alpha$ and $\models \neg\mathbf{L}\alpha \supset \mathbf{L}\neg\mathbf{L}\alpha$).

As far as the operator $\mathbf{O}$ is concerned, its properties are probably best explaine by noting its tight connection to the stable expansions of autoepistemic logic as proposed by Moore [Moo85].

Definition 2 (Moore) *Stable Expansions*
A set of sentences Γ is a **stable expansion** *of a set of sentences A iff Γ satisfies the fixed-point equation:*

$$\Gamma = \{\alpha \mid \alpha \textit{ is basic and } A \cup \{\mathbf{L}\beta \mid \beta \in \Gamma\} \cup \{\neg\mathbf{L}\beta \mid \beta \notin \Gamma\} \models_{\text{taut}} \alpha\},$$

where $\models_{\text{taut}}$ denotes tautological consequence, that is, logical consequence of classical propositional logic.

Definition 3 *Belief Set*
A set of basic sentences is a **belief set** *for a set of worlds M iff $\Gamma = \{\alpha \mid M \models \mathbf{L}\alpha\}$.*

Although belief sets contain non-objective sentences, it turns out that it suffices to consider only the objective ones in order to distinguish between different belief sets.

Lemma 2.1 *Belief sets are uniquely determined by the objective sentences they contain.*

Proof: A proof can be found in [HM84]. ■

The following corollary will be useful in the next section.

Corollary 2.2 *Belief sets are uniquely determined by the objective clauses they contain.*

Proof: Follows directly from the fact that beliefs have equivalent conjunctive normal forms and that $\models \mathbf{L}(\bigwedge \alpha_i) \equiv \bigwedge \mathbf{L}\alpha_i$. ■

The next theorem demonstrates that OL subsumes Moore's autoepistemic logic provided we restrict ourselves to a finite set of premises (understood conjunctively).

Theorem 1 (Levesque) *For any basic α and any set of worlds M, $M \models \mathbf{O}\alpha$ iff the belief set of α is a stable expansion of $\{\alpha\}$.*[6]

3 The Logic OL^a

In this section we extend OL in a way that allows us to express the fact "this is all I know *about* a finite set of atomic propositions π". For that purpose, we add an infinite number of modal operators $\mathbf{O}\langle\pi\rangle$ for each finite set of atoms π to the language. If we refer to a set extensionally, we usually leave out the curly brackets. For example, we write $\mathbf{O}\langle p, q\rangle$ instead of $\mathbf{O}\langle\{p, q\}\rangle$.

Our task is now to define the semantics of $\mathbf{O}\langle\pi\rangle\alpha$, that is, given a set of worlds M, when is it the case that α is all M knows about the atoms in $\pi\Gamma$ We approach this question by first defining a set of worlds $M|_\pi$ that has the same knowledge about π as M but knows nothing else. In a sense, $M|_\pi$ is obtained from M by forgetting everything that is irrelevant to π. With that construction, it seems that believing only α about a subject π at a set of worlds M reduces to believing only α (without any further qualifications) at $M|_\pi$. It turns out, however, that this definition sometimes has the effect that one believes only α about π without actually believing α, which seems counterintuitive.[7] This problem is caused, roughly, by the fact that an introspective agent who forgets not only loses knowledge but also gains knowledge, namely about his or her increased ignorance. In order to circumvent this problem, we simply add the restriction that α must be believed to the definition of $\mathbf{O}\langle\pi\rangle\alpha$. We now turn to the formal definitions.

$M|_\pi$ is defined as the set of all worlds that satisfy precisely the known objective sentences about π. The question is, of course, how to get at just those known sentences about π. First, by Corollary 2.2, it suffices to consider only the known *clauses* instead of arbitrary sentences. Furthermore, we only need to look at clauses that are minimal in the sense that they do not already follow from other known clauses:

[6] Levesque also shows that we can drop the condition that sentences are basic if we generalize the notion of stable expansions appropriately. Also, for reasons independent of this theorem, Levesque restricts himself to so-called maximal sets of worlds, a complication which we simply ignore here.

[7] An example is given in Section 3.1.

Definition 4 *M-minimality*
*Given a set of worlds M, a clause c is called M-**minimal** iff $M \models \mathbf{L}c$ and for all clauses $c' \subsetneq c$, $M \not\models \mathbf{L}c'$.*

By restricting ourselves to M-minimal clauses, we rule out clauses that mention the subject matter but do not really tell us anything about it. For example, let the subject matter be p and assume all we know is q, that is, $M = \{w \mid w \models q\}$. Then we certainly also know $(p \vee q)$, which is not M-minimal because q is known as well. While $(p \vee q)$ mentions the subject matter p, it does so, in a sense, only accidentally, since it does not convey us what is really known about p, namely nothing. The only M-minimal clause mentioning p is $(p \vee \neg p)$, which gives us the right information.

In general, what is known about a subject matter π at a set of worlds M is captured by those M-minimal clauses that mention any of the atomic propositions in π. The only exception occurs when M is empty. In this case M knows every sentence and has, in particular, contradictory about the subject matter, which is characterized by $\Box$, the only M-minimal clause if M is empty. This leads us to the following definition.

Definition 5 *M-π-minimality*
*Given a set of worlds M and a subject matter π, a clause c is called M-π-**minimal** iff c is M-minimal and, in addition, $c = \Box$ or c contains either p or $\neg p$ for some $p \in \pi$.*

$M|_\pi$ is now simply the set of all worlds that satisfy all M-π-minimal sentences.

Definition 6
Given a set of worlds M and a subject matter π,

$$M|_\pi = \{w \mid w \models c \text{ for all } M\text{-}\pi\text{-minimal clauses } c\}.$$

Given $M|_\pi$, we are now able to characterize the meaning of $\mathbf{O}\langle\pi\rangle$ in terms $\mathbf{O}$ and $\mathbf{L}$:

$$M, w \models \mathbf{O}\langle\pi\rangle\alpha \quad \Longleftrightarrow \quad M|_\pi, w \models \mathbf{O}\alpha \text{ and } M \models \mathbf{L}\alpha.$$

Logical implication and *validity* in OL^a are defined as in OL.

This completes the semantics of the extended logic OL^a.

3.1 Some properties of OL^a

So far, we have not obtained a complete axiomatization of OL^a. In the following, we present examples that demonstrate that the new concept $\mathbf{O}\langle\pi\rangle$ has indeed reasonable properties. (The proofs are deferred to the end of this subsection.) A more concise picture of what $\mathbf{O}\langle\pi\rangle$ is all about will emerge in the next section, where we show a strong connection between only-knowing-about and the ATMS.

Let α and β be sentences and p and q distinct atoms.

1. $\models \mathbf{O}\langle\pi\rangle\alpha \supset \mathbf{L}\alpha$,
   that is, if α is all I believe about π then I surely believe α. However, ...

2. $\not\models \mathbf{O}\langle\pi\rangle\alpha \supset \mathbf{O}\alpha$ (unless $\models \neg\mathbf{O}\langle\pi\rangle\alpha$),
   since I could also believe q for any atom not in π or α. On the other hand, if π contains all the atoms occurring in α, only-knowing and only-knowing-about coincide, that is, $\models \mathbf{O}\langle\pi\rangle\alpha \equiv \mathbf{O}\alpha$.

3. If $\models \alpha \equiv \beta$, then $\models \mathbf{O}\langle\pi\rangle\alpha \equiv \mathbf{O}\langle\pi\rangle\beta$.
   If all I believe about π is α, then the syntactic form of α is immaterial.

4. $\models \mathbf{O}\langle p\rangle(p \vee q) \supset (\neg\mathbf{L}p \wedge \neg\mathbf{L}q)$.
   This is because if I knew either p or q, then $p \vee q$ would be contingent information and thus, in either case, would not capture what I really know about p. Similarly,

5. $\models \mathbf{O}\langle p\rangle(p \vee q) \supset (\neg\mathbf{L}\neg p \wedge \neg\mathbf{L}\neg q)$.
   For if I knew $\neg p$, I would also know q because I know $(p \vee q)$. Thus $(p \vee q)$ would once again not capture what I really know about p. (Similarly, if I knew $\neg q$.)

6. $\models \mathbf{O}\langle p, q\rangle p \supset \mathbf{O}\langle q\rangle(q \vee \neg q)$.
   If all I know about p and q is p, then I know nothing about q.

   The following example treats a case where it is absurd to say this is all I know about p.

7. $\models \neg\mathbf{O}\langle p\rangle q$.
   In other words, something totally independent of p cannot be all I know about p. For example, it does not make sense to say that all I know about Tweety is that roses are red.

8. As Theorem 1 shows, only-knowing captures autoepistemic reasoning. The following example indicates that the weaker assumption of only-knowing-about in fact suffices for this purpose.

   Let p denote the proposition *Tweety flies*. Given the assumption that Tweety flies unless known otherwise, autoepistemic reasoning lets us conclude that Tweety flies. Formally, $\models \mathbf{O}(\neg\mathbf{L}\neg p \supset p) \supset \mathbf{L}p$. Later, if we discover that Tweety indeed does not fly, we retract our previous conclusion, that is, $\models \mathbf{O}(\neg p \wedge (\neg\mathbf{L}\neg p \supset p)) \supset \neg\mathbf{L}p$. However, it is intuitively clear that we need not require that the assumption is all we know to get the desired conclusion, but that the weaker requirement that this is all we know *about Tweety* suffices. And indeed, our formalization gives us just that. (The subject matter *about Tweety* in this case is simply $\{p\}$.)

   $\models \mathbf{O}\langle p\rangle(\neg\mathbf{L}\neg p \supset p) \supset \mathbf{L}p$.
   $\models \mathbf{O}\langle p\rangle(\neg p \wedge (\neg\mathbf{L}\neg p \supset p)) \supset \neg\mathbf{L}p$.

9. The following example concerns the case of sentences with multiple stable expansions. Again, the weaker requirement of only-knowing-about suffices to make the same distinctions as regular autoepistemic logic.

   $\models \mathbf{O}\langle p\rangle(\mathbf{L}p \supset p) \equiv [\mathbf{O}\langle p\rangle p \vee \mathbf{O}\langle p\rangle(p \vee \neg p)]$.
   In other words, believing only $\mathbf{L}p \supset p$ about p is the same as believing either only p or nothing about p.

Proofs:

1. $\models \mathbf{O}\langle\pi\rangle\alpha \supset \mathbf{L}\alpha$.

   Follows immediately from the definition of $\mathbf{O}\langle\pi\rangle$.

   It is worth noting that the property does not hold in general, if we omit the condition '$M\models\mathbf{L}\alpha$' from the definition of $\mathbf{O}\langle\pi\rangle$. Here is an example. Let $M^* = \{w \mid w\models q\}$. Then M^* knows absolutely nothing about p, that is, $M^*\models\mathbf{O}\langle p\rangle(p\vee\neg p)$ or, equivalently, $M^*\models\mathbf{O}\langle p\rangle\neg\mathbf{L}p$ because $M^*|_p = M_0$. Note that $M^*|_p\models\mathbf{O}\neg\mathbf{L}q$, since it knows nothing about q. However, $M^*\not\models\mathbf{L}\neg\mathbf{L}q$.

2. $\not\models\mathbf{O}\langle\pi\rangle\alpha \supset \mathbf{O}\alpha$ (unless $\models\neg\mathbf{O}\langle\pi\rangle\alpha$).

   Let $\alpha = p$, $\pi = \{p\}$, and let $M = \{w \mid w\models p\wedge q\}$. Since the only M-minimal clause that mentions p is p itself, $M|_\pi = \{w \mid w\models p\}$. Since $M\models\mathbf{L}p$ and $M|_\pi\models\mathbf{O}p$, we obtain $M\models\mathbf{O}\langle\pi\rangle p$, but $M\not\models\mathbf{O}p$. In fact, $M\models\mathbf{O}(p\wedge q)$.

3. If $\models\alpha\equiv\beta$, then $\models\mathbf{O}\langle\pi\rangle\alpha\equiv\mathbf{O}\langle\pi\rangle\beta$.

   Let $\models\alpha\equiv\beta$ and let M be a set of worlds such that $M\models\mathbf{O}\langle\pi\rangle\alpha$. We need to show that $M\models\mathbf{O}\langle\pi\rangle\beta$. (The reverse direction is completely symmetric and is omitted.)

   By assumption, $M\models\mathbf{L}\alpha$, that is, for all $w\in M$, $M,w\models\alpha$. Since $\models\alpha\equiv\beta$, we obtain that for all $w\in M$, $M,w\models\beta$, which implies that $M\models\mathbf{L}\beta$.

   Since, by assumption, $M|_\pi\models\mathbf{O}\alpha$, we obtain that for all worlds w, $w\in M|_\pi$ iff $M|_\pi,w\models\alpha$. Since $\models\alpha\equiv\beta$, this is the same as for all w, $w\in M|_\pi$ iff $M|_\pi,w\models\beta$ and, therefore, $M|_\pi\models\mathbf{O}\beta$.

   Since $M\models\mathbf{L}\beta$ and $M|_\pi\models\mathbf{O}\beta$, $M\models\mathbf{O}\langle\pi\rangle\beta$ follows.

4. $\models\mathbf{O}\langle p\rangle(p\vee q)\supset(\neg\mathbf{L}p\wedge\neg\mathbf{L}q)$.

   Let $M\models\mathbf{O}\langle p\rangle(p\vee q)$. Then $M|_p\models\mathbf{O}(p\vee q)$, that is, $M|_p = \{w \mid w\models(p\vee q)\}$. Assume that $M\models\mathbf{L}p$. Then $M|_p\models\mathbf{L}p$, since p is M-minimal. However, $M|_p$ contains a world w^* such that $w^*\models q$ yet $w^*\not\models p$, a contradiction.

   Now assume that $M\models\mathbf{L}q$. Then none of the M-minimal clauses that mention p contains q as a literal. (In particular, $(p\vee q)$ is no longer M-minimal.) As noted before, $M|_p$ contains a world w^* such that $w^*\models q$ yet $w^*\not\models p$. By definition of $M|_p$, $w^*\models c$ for all M-minimal clauses c that mention p. Let w^{**} be exactly like w^* except that $w^{**}\models\neg q$. Note that $w^{**}\models c$ for all M-minimal clauses c that mention p, because none of those c contains the literal q. Thus $w^{**}\in M|_p$, but $w^{**}\not\models(p\vee q)$, a contradiction.

5. $\models\mathbf{O}\langle p\rangle(p\vee q)\supset(\neg\mathbf{L}\neg p\wedge\neg\mathbf{L}\neg q)$.

   Let $M\models\mathbf{O}\langle p\rangle(p\vee q)$. Then $M|_p\models\mathbf{O}(p\vee q)$, that is, $M|_p = \{w \mid w\models(p\vee q)\}$. Assume $M\models\mathbf{L}\neg p$. Then, since $M\models\mathbf{L}(p\vee q)$, $M\models\mathbf{L}q$ follows, contradicting the previous property. Similarly, if we assume that $M\models\mathbf{L}\neg q$, then $M\models\mathbf{L}p$ follows, a contradiction.

6. $\models\mathbf{O}\langle p,q\rangle p\supset\mathbf{O}\langle q\rangle(q\vee\neg q)$.

   Let $\pi = \{p,q\}$ and $M\models\mathbf{O}\langle\pi\rangle p$. Assume $M\not\models\mathbf{O}\langle q\rangle(q\vee\neg q)$. Thus $M|_q\not\models\mathbf{O}(q\vee\neg q)$ (since $M\models\mathbf{L}(q\vee\neg q)$) and, hence, $M|_q\neq M_0$ (the set of all worlds). In particular, there is a clause c^* which is M-minimal and mentions q and which is not a tautology. Note also that c^* does not contain the literal p because p itself is M-minimal. By the definition of $M|_\pi$, $M|_\pi \models \mathbf{L}c^*$. But $M|_\pi = \{w \mid w \models p\}$. Thus there is a world $w^* \in M|_\pi$ such that $w^* \not\models c^*$ (since c^* does not contain the literal p), a contradiction.

7. $\models \neg \mathbf{O}\langle p \rangle q$.

   Assume, to the contrary, that there is an M such that $M \models \mathbf{O}\langle p \rangle q$. Then $M|_p \models \mathbf{O}q$, that is, $M|_p = \{w \mid w \models q\}$. Since $M \models \mathbf{L}q$, q is M-minimal and no M-minimal clause that mentions p contains q as a literal. Thus there exists a w^* that satisfies all M-minimal clauses mentioning p and $w^* \not\models q$. But $w^* \in M|_p$, a contradiction.

8. $\models \mathbf{O}\langle p \rangle (\neg \mathbf{L} \neg p \supset p) \supset \mathbf{L}p$.

   Let $M \models \mathbf{O}\langle p \rangle (\neg \mathbf{L} \neg p \supset p)$. Then $M|_p \models \mathbf{O}(\neg \mathbf{L} \neg p \supset p)$ and thus $M|_p = \{w \mid w \models p\}$, that is, $M|_p \models \mathbf{L}p$. Since $M \subseteq M|_p$, $M \models \mathbf{L}p$ holds as well.

   $\models \mathbf{O}\langle p \rangle (\neg p \wedge (\neg \mathbf{L} \neg p \supset p)) \supset \neg \mathbf{L}p$.

   Similar to the previous case, if we assume that $M \models \mathbf{O}\langle p \rangle (\neg p \wedge (\neg \mathbf{L} \neg p \supset p))$, then $M|_p \models \mathbf{O}(\neg p \wedge (\neg \mathbf{L} \neg p \supset p))$ and thus $M|_p = \{w \mid w \models \neg p\}$, which implies that $M \models \mathbf{L} \neg p$ and $M \models \neg \mathbf{L}p$.

9. $\models \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p) \equiv [\mathbf{O}\langle p \rangle p \vee \mathbf{O}\langle p \rangle (p \vee \neg p)]$.

   First we prove that $\models \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p) \supset [\mathbf{O}\langle p \rangle p \vee \mathbf{O}\langle p \rangle (p \vee \neg p)]$. Let $M \models \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p)$. Then $M|_p \models \mathbf{O}(\mathbf{L}p \supset p)$. It is easy to see that there are only two cases: (a) $M|_p = \{w \mid w \models p\}$ and (b) $M|_p = M_0$. In case (a) we obtain $M \models \mathbf{L}p$, $M|_p \models \mathbf{O}p$, and, hence, $M \models \mathbf{O}\langle p \rangle p$. In case (b), $M \models \mathbf{L}(p \vee \neg p)$, $M|_p \models \mathbf{O}(p \vee \neg p)$, and, hence, $M \models \mathbf{O}\langle p \rangle (p \vee \neg p)$.

   To prove $\models [\mathbf{O}\langle p \rangle p \vee \mathbf{O}\langle p \rangle (p \vee \neg p)] \supset \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p)$, let us first assume that $M \models \mathbf{O}\langle p \rangle p$. Then $M \models \mathbf{L}p$ and $M|_p \models \mathbf{O}p$. Therefore, $M \models \mathbf{L}(\mathbf{L}p \supset p)$ and $M|_p \models \mathbf{O}(\mathbf{L}p \supset p)$, which implies $M \models \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p)$. Now let us assume that $M \models \mathbf{O}\langle p \rangle (p \vee \neg p)$. Then $M \models \mathbf{L} \neg \mathbf{L}p$ and, hence, $M \models \mathbf{L}(\mathbf{L}p \supset p)$. Also, since $M|_p \models \mathbf{O}(p \vee \neg p)$, $M|_p \models \mathbf{O}(\mathbf{L}p \supset p)$ holds as well. Again, $M \models \mathbf{O}\langle p \rangle (\mathbf{L}p \supset p)$ follows. ■

3.2 Levesque's Proposal

As mentioned before, Levesque suggests a definition of $\mathbf{O}\langle \pi \rangle$ in [Lev90], yet without analyzing it in any depth. We now briefly discuss Levesque's proposal and point out some of its problems.

In our notation and restricted to the propositional case, Levesque's definition goes as follows:

$$M \models \mathbf{O}\langle \pi \rangle \alpha \iff \text{for all } w,\ w \in M \text{ iff } M, w \models \alpha \text{ and there is a } w' \in M \text{ such that for all atoms } p, \text{ if } w(p) \neq w'(p), \text{ then } p \in \pi.$$

The extra condition (compared to the definition of $\mathbf{O}$) amounts to saying that w must be an element of M unless it violates knowledge about something other than π.

At first glance, this definition looks quite appealing because of its simplicity. It also shares many of the properties with our definition of $\mathbf{O}\langle \pi \rangle$. Unfortunately, it also has some rather counterintuitive properties. Here are two examples. In the following, we abbreviate the condition "... *and there is a* $w' \in M$ *such that for all atoms* p*, if* $w(p) \neq w'(p)$*, then* $p \in \pi$" by $(*)$.

1. Let $M = \{w \mid w \models q\}$. Intuitively, M knows nothing about p. However, according to Levesque's definition, $M \models \mathbf{O}\langle p \rangle q$, that is, M claims to know something totally irrelevant about p.

   Proof: First, let $w \in M$. Obviously $w \models (p \vee q)$. The condition $(*)$ is satisfied simply by choosing $w' = w$. If $w \notin M$, then $w \not\models q$ and we are done.

2. Let M be as before. Then $M \models \mathbf{O}\langle p \rangle (p \supset q)$. This seems even more misleading, since M's alleged knowledge about p seems at least plausible.

   Proof: For any $w \in M$, $w \models q$ and w satisfies $(*)$. Conversely, let $w \notin M$. We have to show that if $w \models (p \supset q)$, then $(*)$ is violated. Thus let us assume that $w \models (p \supset q)$. Then $w \not\models p$ and $w \not\models q$. Therefore, any $w' \in M$ is such that $w(q) \neq w'(q)$, but $q \notin \{p\}$.

While it seems possible to filter out occurrences of example (1) by adding a suitable restriction that forces α to be relevant to π, case (2) cannot be resolved by looking at α and π alone.

4 Computing all that is known about some subject matter

Having defined what it means for α to be all an agent knows about a certain subject matter π, it seems natural to ask how to compute α from the agent's knowledge base (KB). In this section we will provide an answer for the special yet important case of *objective* KBs, that is, no modal operators are allowed in the KB. This case is particularly interesting because computing all that is known about π reduces to abductive reasoning and, in particular, to computing *explanations* in the sense of an ATMS [deK86] for the atoms in π and their negations. Thus we are able to relate only-knowing-about to more familiar notions in knowledge representation.

Following [Lev90], a sentence α is said to explain a sentence β just in case $\alpha \supset \beta$ is believed and $\neg\alpha$ is not believed. If $\mathbf{L}$ is our model of belief and the beliefs are represented by an objective KB,[8] Levesques has shown that his definition coincides with previous definitions of abduction such as Poole's [Poo88], where α explains β iff $\text{KB} \cup \{\alpha\} \models \beta$ and $\text{KB} \cup \{\alpha\}$ is consistent.

An ATMS then computes the set of all simplest explanations[9] for a given KB and β.

As an example, let us consider a simple minded medical KB which only knows that hepatitis causes jaundice and that the patient has a fever, represented as

$$\text{KB} = \{(hepatitis \supset jaundice), fever\}.$$

An ATMS, asked for the explanations of *jaundice*, returns both *hepatitis* and the trivial explanation *jaundice* itself, denoted as $\mathtt{atms}[\text{KB}, jaundice] = \{jaundice, hepatitis\}$. Notice that $hepatitis \supset jaundice$ is all that is known about jaundice according to our definition. We can reconstruct this information systematically using the output from the ATMS as

$$\alpha = \bigwedge_{c \in \mathtt{atms}[\text{KB}, jaundice]} (c \supset jaundice)$$

[8] The beliefs represented by KB are characterized by the set of all worlds that satisfy KB.

[9] Roughly, an explanation α is simpler than β if the literals in α are contained in β.

It is easy to verify that

$$\models \mathbf{O}\text{KB} \supset \mathbf{O}\langle jaundice \rangle \alpha$$

because $\alpha \equiv (jaundice \supset jaundice) \wedge (hepatitis \supset jaundice) \equiv (hepatitis \supset jaundice)$. In the following, we examine the general case of computing all that is known about a subject matter using an ATMS (Theorem 2).

To simplify matters, we assume that the subject matter π consists of a single atom p. All of the following results generalize in a straightforward way to arbitrary sets due to the following lemma:

Lemma 4.1 *Let π_1 and π_2 be two sets of atoms and α and β objective sentences. Then $\models \mathbf{O}\langle \pi_1 \rangle \alpha \wedge \mathbf{O}\langle \pi_2 \rangle \beta \supset \mathbf{O}\langle \pi_1 \cup \pi_2 \rangle (\alpha \wedge \beta)$.*

Proof : Let M be a set of worlds such that $M \models \mathbf{O}\langle \pi_1 \rangle \alpha \wedge \mathbf{O}\langle \pi_2 \rangle \beta$. By the semantics of $\mathbf{O}\langle \pi_1 \rangle$ and $\mathbf{O}\langle \pi_2 \rangle$, $M \models \mathbf{L}\alpha$ and $M \models \mathbf{L}\beta$ and hence $M \models \mathbf{L}(\alpha \wedge \beta)$. Let $\pi = \pi_1 \cup \pi_2$. We still need to prove that $M|_\pi \models \mathbf{O}(\alpha \wedge \beta)$.

First we show that $M|_\pi = M|_{\pi_1} \cap M|_{\pi_2}$. If $M = \{\}$ then $M|_\pi = M|_{\pi_1} = M|_{\pi_2} = \{\}$ and we are done. Now assume that M is not empty. Then $M|_\pi$, $M|_{\pi_1}$, and $M|_{\pi_2}$ are not empty either. To show that $M|_\pi \models \mathbf{O}(\alpha \wedge \beta)$, let w be any world. $w \in M|_\pi$ iff $w \models c$ for all M-minimal clauses c that mention some $p \in \pi$ iff $w \models c$ for all M-minimal clauses c that mention some $p \in \pi_1$ and $w \models c$ for all M-minimal clauses c that mention some $p \in \pi_2$ iff $w \in M|_{\pi_1} \cap M|_{\pi_2}$.

By assumption, $M|_{\pi_1} \models \mathbf{O}\alpha$ and $M|_{\pi_2} \models \mathbf{O}\beta$. Since α and β are objective, $M|_{\pi_1} = \{w \mid w \models \alpha\}$ and $M|_{\pi_2} = \{w \mid w \models \beta\}$. Therefore $M|_\pi = \{w \mid w \models \alpha\} \cap \{w \mid w \models \beta\}$. From this it is straightforward to show that $M|_\pi \models \mathbf{O}(\alpha \wedge \beta)$. ∎

Notation: For any clause c, let $\overline{c}$ denote the complement of c, that is, the conjunction of the complements of the literals contained in c. Also, from now on KB denotes a finite set of clauses. Whenever the formalism requires a sentence instead of a set, we write KB as well and mean the conjunction of all the clauses in the knowledge base. Let $\Re[\text{KB}] = \{w | w \models \text{KB}\}$, that is $\Re[\text{KB}]$ represents the beliefs that follow from KB.

The following definitions are mainly adapted from [Lev89].

First we need to define what it means for a sentence to be a simplest (or minimal) explanation.

Definition 7 *Simplicity*
The set of literals contained in an objective sentence is defined as follows:
$\text{LITS}(\Box) = \{\}; \text{LITS}(p) = \{p\}; \text{LITS}(\neg\alpha) = \{\overline{l} \mid l \in \text{LITS}(\alpha)\};$
$\text{LITS}(\alpha \wedge \beta) = \text{LITS}(\alpha) \cup \text{LITS}(\beta).$

α is simpler than β ($\alpha \prec \beta$) iff $\text{LITS}(\alpha) \subsetneq \text{LITS}(\beta)$.

Definition 8 *Explanations and minimal explanations*
Let α and β be objective sentences. Then

1. α $\text{expl}_{\mathbf{L}}$ β *wrt.* KB *iff* $\Re[\text{KB}] \models \mathbf{L}(\alpha \supset \beta) \wedge \neg\mathbf{L}\neg\alpha$.

2. α $\text{min\_expl}_{\mathbf{L}}$ β *wrt.* KB *iff* α $\text{expl}_{\mathbf{L}}$ β *wrt.* KB *and for no* $\alpha^* \prec \alpha$, α^* $\text{expl}_{\mathbf{L}}$ β *wrt.* KB.

We now turn to Reiter and deKleer's formalization of an ATMS [RdK87].

Definition 9 *An ATMS*

1. $\text{IMPS}[\text{KB}] = \{c \mid c \textit{ is a clause and } \text{KB} \models c\}$.

2. $\mu\text{IMPS}[\text{KB}] = \{c \mid c \in \text{IMPS}[\text{KB}] \textit{ and for no } c' \subsetneq c, c' \in \text{IMPS}[\text{KB}]\}$.[10]

3. *Let c be a clause. Then*
   $\texttt{atms}[\text{KB}, c] = \{(q_1 \wedge \ldots \wedge q_k) \mid k \geq 0 \text{ and } (\overline{q_1} \vee \ldots \vee \overline{q_k} \vee c) \in \mu\text{IMPS}[\text{KB}]\}$.[11]

Note that `atms` returns explanations in a special syntactic form, namely as conjunctions of literals. Levesque's definition of simplest explanations, on the other hand, does not worry about how an explanation is represented. Nevertheless, as shown by Levesque, the two notions are equivalent at the level of propositions,[12] that is, the propositions expressed by $\texttt{atms}[\text{KB}, c]$ are the same as the propositions expressed by the simplest explanations of c with respect to KB. Given Definition 9, we are now able to prove that all that is known about p by KB is completely characterized by $\texttt{atms}[\text{KB}, p]$ together with $\texttt{atms}[\text{KB}, \neg p]$. First, we note a straightforward connection between IMPS[KB] and the notion of $\Re[\text{KB}]$-minimality (Definition 4).

Lemma 4.2 *For any clause c, $c \in \mu\text{IMPS}[\text{KB}]$ iff c is $\Re[\text{KB}]$-minimal.*

Proof : Note that $\Re[\text{KB}] \models \mathbf{L}\alpha$ iff $\text{KB} \models \alpha$ for all objective sentences α. The lemma follows now immediately from the definitions of $\Re[\text{KB}]$-minimality and $\mu\text{IMPS}[\text{KB}]$. ■

Theorem 2 *Let p be an atom, $\Gamma_p = \texttt{atms}[\text{KB}, p]$, and $\Gamma_{\overline{p}} = \texttt{atms}[\text{KB}, \neg p]$. Then*

$$\models \mathbf{O}\text{KB} \supset \mathbf{O}\langle p\rangle[\bigwedge_{c \in \Gamma_p} (c \supset p) \wedge \bigwedge_{c \in \Gamma_{\overline{p}}} (c \supset \neg p)].$$

Proof :

Let $\alpha = \bigwedge_{c \in \Gamma_p}(c \supset p) \wedge \bigwedge_{c \in \Gamma_{\overline{p}}}(c \supset \neg p)$. (Note that $(c \supset p)$ and $(c \supset \neg p)$ really are clauses, since c is a conjunction of literals.) First, it is not hard to show that both Γ_p and $\Gamma_{\overline{p}}$ are finite so that α is in fact a well-defined objective sentence. Let $M = \{w \mid w \models \alpha\}$. Then $M = \{w \mid w \models (c \supset p) \text{ for all } c \in \Gamma_p\} \cap \{w \mid w \models (c \supset \neg p) \text{ for all } c \in \Gamma_{\overline{p}}\}$. By Lemma 4.2, $\Gamma_p = \{c \mid (c \supset p) \textit{ is } \Re[\text{KB}]\text{-minimal}\}$ and $\Gamma_{\overline{p}} = \{c \mid (c \supset \neg p) \textit{ is } \Re[\text{KB}]\text{-minimal}\}$. Note that, by Definition 6,

$$\begin{aligned} \Re[\text{KB}]|_p = & \{w \mid w \models (c \supset p) \text{ for all } \Re[\text{KB}]\text{-minimal } (c \supset p)\} \quad \cap \\ & \{w \mid w \models (c \supset \neg p) \text{ for all } \Re[\text{KB}]\text{-minimal } (c \supset \neg p)\} \end{aligned}$$

Thus $\Re[\text{KB}]|_p = M$ and, hence, $\Re[\text{KB}]|_p \models \mathbf{O}\alpha$. Obviously, $M \models \mathbf{L}\alpha$ holds as well and, therefore, $M \models \mathbf{O}\langle p\rangle\alpha$.

[Note that in the case where KB is consistent and $\text{KB} \models p$, $\texttt{atms}[\text{KB}, p] = \{\neg\Box\}$, in which case $\alpha \equiv (\neg\Box \supset p) \equiv p$ (similarly for $\text{KB} \models \neg p$). If KB is inconsistent, then $\texttt{atms}[\text{KB}, p] = \{\}$. Thus α is the empty clause, represented as $\Box$.] ■

[10] In the terminology of Reiter and deKleer, μIMPS[KB] contains the *prime implicants* of KB.
[11] Actually, given deKleer's original formulation of an ATMS [deK86], this came out as a theorem in [RdK87].
[12] The proposition expressed by a sentence α can be thought of as the set of worlds that satisfy α.

5 The Logic OB^a

While the previous section demonstrates how only-knowing-about can in principle be computed using an ATMS, the result also entails that the computational cost is considerable since the ATMS is inherently intractable [SL90].

The source of this complexity lies in the very powerful model of belief that we adopted, which requires belief to be closed under logical implication ($\models \mathbf{L}\alpha \wedge \mathbf{L}(\alpha \supset \beta) \supset \mathbf{L}\beta$). In particular, this means that even computing the objective beliefs of $\Re[\text{KB}]$ is intractable. Levesque has called this type of belief *implicit belief* and contrasts it with *explicit belief*, which should take into account a reasoner's limited resources [Lev84]. In [Lev89], Levesque suggests a particular form of explicit belief,[13] which does not only yield a tractable deductive reasoner, but which lends itself to the definition of a tractable form of abduction or limited ATMS. Given this result, we are able to show that an agent can efficiently compute all that he or she explicitly knows about some subject matter.

First we modify the logic OL^a in order to incorporate the notion of explicit belief along the lines of [Lev89].

For this purpose, we replace the operator $\mathbf{L}$ with a new operator $\mathbf{B}$ (for explicit belief). $\mathbf{B}$ and $\mathbf{O}$ in this new logic are variants of the same operators in [LL88]. Also $\mathbf{B}$ extends Levesque's notion of explicit belief in [Lev89] (which, in turn, is a variant of Levesque's logic of explicit belief in [Lev84]) by allowing arbitrary nestings of modal operators.

The semantics of this new logic OB^a is defined just like the semantics of OL^a except that $\mathbf{B}$ and $\mathbf{O}$ are interpreted with respect to so-called *situations* instead of worlds and $\mathbf{O}\langle\pi\rangle$ is defined in terms of $\mathbf{B}$ and $\mathbf{O}$. While worlds assign a truth value to every atomic proposition, situations, in contrast, assign truth values to all the *literals* with the restriction that either p or $\neg p$ is assigned `true`. This gives us the following important difference between worlds and situations: while a literal and its complement always have complementary truth values at a world, they may both have the value $\mathbf{t}$ at a situation.

Definition 10 *Situations*
A situation s is a function s : Literals $\longrightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $s(\Box) = \mathbf{f}$ and for every atom p, either $s(p) = \mathbf{t}$ or $s(\neg p) = \mathbf{t}$.

The definitions for M-minimality, M-π-minimality, and $M|_\pi$ carry over from implicit belief to explicit belief in the obvious way. Note that those definitions rely on the fact that belief sets are uniquely determined by the objective clauses they contain. Lemma 5.1 below verifies that this is indeed the case for explicit belief as well.

Definition 11 *M-minimality*
*Given a set of situations M, a clause c is called M-**minimal** iff $M \models \mathbf{B}c$ and for all clauses $c' \subsetneq c$, $M \not\models \mathbf{B}c'$.*

Definition 12 *M-π-minimality*
*Given a set of situations M and a subject matter π, a clause c is called M-π-**minimal** iff c is M-minimal and, in addition, $c = \Box$ or c contains either p or $\neg p$ for some $p \in \pi$.*

[13] This is a variant of the type of explicit belief defined in [Lev84].

Definition 13
Given a set of situations M and a subject matter π,

$$M|_\pi = \{s \mid s \text{ is a situation and } s \models c \text{ for all } M\text{-}\pi\text{-minimal clauses } c\}.$$

Since situations assign independent truth values to literals and their complements, the semantic rules for arbitrary sentences must define truth support for sentences and their negations. Let p be an atom, α and β arbitrary sentences, s a situation, and M a set of situations. $M|_\pi$ is defined as in Section 3 except that worlds are replaced by situations.

$M, s \models p \iff s(p) = \mathbf{t}$
$M, s \models \neg p \iff s(\neg p) = \mathbf{t}$
$M, s \models \neg\neg\alpha \iff M, s \models \alpha$
$M, s \models \alpha \vee \beta \iff M, s \models \alpha$ or $M, s \models \beta$
$M, s \models \neg(\alpha \vee \beta) \iff M, s \models \neg\alpha$ and $M, s \models \neg\beta$
$M, s \models \mathbf{B}\alpha \iff$ for all s', if $s' \in M$ then $M, s' \models \alpha$
$M, s \models \mathbf{O}\alpha \iff$ for all s', $s' \in M$ iff $M, s' \models \alpha$
$M, s \models \mathbf{O}\langle\pi\rangle\alpha \iff M|_\pi, s \models \mathbf{O}\alpha$ and $M, s \models \mathbf{B}\alpha$
Let Φ be any of the operators $\mathbf{B}$,$\mathbf{O}$, or $\mathbf{O}\langle\pi\rangle$.
$M, s \models \neg\Phi\alpha \iff M, s \not\models \Phi\alpha$

Since we are interested in situations only as far as explicit belief is concerned, we define truth and logical implication, as usual, with respect to worlds only. Let w be a world and M a set of situations. A sentence α is said to be true at w and M just in case $M, w \models \alpha$. For a set of sentences Γ, $\Gamma \models \alpha$ iff for all worlds w and sets of situations M, if $M, w \models \gamma$ for all $\gamma \in \Gamma$, then $M, w \models \alpha$. Validity ($\models \alpha$) is defined as usual as $\{\} \models \alpha$.

5.1 Some properties of OB^a

As said before, explicit belief ($\mathbf{B}$) and explicitly only-believing ($\mathbf{O}$) are variants of similar notions in [LL88] (and also extensions of [Lev84, Lev89]). Thus we will not discuss the properties of $\mathbf{B}$ and $\mathbf{O}$ in detail here. What makes implicit belief interesting for our purposes is that it allows for tractable reasoning in the sense that, if we have an objective KB and an objective α, both in conjunctive normal form, then deciding whether $\mathbf{B}\alpha$ is logically implied by $\mathbf{B}$KB is tractable [Lev89]. The algorithm is very simple and relies on the following fact. $\mathbf{B}\alpha$ is logically implied by $\mathbf{B}$KB iff for every clause c in α one of the following properties holds:

1. c is tautologous, that is, it contains complementary literals.
2. There is a clause c' in KB such that $c' \subseteq c$.

The corresponding algorithm, of course, can be computed in time $O(|\text{KB}|, |\alpha|)$.[14] Notice that *modus ponens* is ruled out as a valid inference rule, that is, an agent who believes p and $p \supset q$ does not necessarily believe q.

Let us now turn to $\mathbf{O}\langle\pi\rangle$ in the context of explicit belief. Our approach to defining the meaning of $\mathbf{O}\langle\pi\rangle$ presumes that the beliefs at a set of situations are uniquely determined by

[14] Apart from testing for tautologous clauses, the algorithm in fact computes *tautological entailment*, a form of relevance logic [AB75, Dunn 1976].

the objective clauses they contain, just as in the case of implicit belief. The following lemma verifies that this is indeed the case.

Lemma 5.1 *The beliefs at a set of situations M are uniquely determined by the clauses that are believed at M.*

Proof : It suffices to prove that (1) the beliefs at M are uniquely determined by the objective sentences that are believed at M, (2) every belief can be transformed into an equivalent conjunctive normal form, and (3) $\models \mathbf{B} \bigwedge \alpha_i \equiv \bigwedge \mathbf{B}\alpha_i$. Proofs of these properties in the case of the logic of [LL88] carry over to OB^a in a straightforward way. ■

In Section 3.1, we proved properties of $\mathbf{O}\langle\pi\rangle$ in the case of implicit belief. Except for property 3, all of those hold in the case of explicit belief as well. In fact, the proofs carry over simply by replacing *worlds* by situations and $\mathbf{L}$ by $\mathbf{B}$.

To see that property 3 no longer holds, let $\alpha = p \wedge (p \supset q)$ and $\beta = p \wedge q$. Obviously, $\models \alpha \equiv \beta$. Let $M = \{s \mid \text{s is a situation and } s \models \alpha\}$. It is easy to see that $M \models \mathbf{O}\langle p, q\rangle\alpha$, but M contains a situations s^* such that $s \models p$, $s \models \neg p$, and $s \not\models q$. Thus $M \not\models \mathbf{B}\beta$ and, hence, $M \not\models \mathbf{O}\langle p, q\rangle\beta$.

The following weakened form of property 3 holds, however, in the case of explicit belief:

3' If for any situation s and any set of situations M, $M, s \models \alpha$ iff $M, s \models \beta$,
then $\models \mathbf{O}\langle\pi\rangle\alpha \equiv \mathbf{O}\langle\pi\rangle\beta$.
(The proof is very similar to the proof of 3.)

5.2 Computing all that is known explicitly about some subject matter

In order to compute what an objective KB knows about some subject matter based on explicit belief, we use Levesque's definition of a limited form of abductive explanation, which is obtained from Definition 8 simply be replacing $\mathbf{L}$ with $\mathbf{B}$. Also, we need to characterize the beliefs of an objective KB now in terms of situations rather than worlds. Therefore, let

$$\Re[\text{KB}] = \{s \mid s \text{ is a } \textbf{situation} \text{ and } s \models \text{KB}\}.$$

Definition 14 (Levesque) *Explanations and minimal explanations for explicit belief*
Let α and β be objective sentences. Then

1. α expl$_{\mathbf{B}}$ β *wrt.* KB *iff* $\Re[\text{KB}] \models \mathbf{B}(\alpha \supset \beta) \wedge \neg\mathbf{B}\neg\alpha$.

2. α min_expl$_{\mathbf{B}}$ β *wrt.* KB *iff* α expl$_{\mathbf{B}}$ β *wrt.* KB *and for no* $\alpha^* \prec \alpha$, α^* expl$_{\mathbf{B}}$ β *wrt.* KB.

Similarly, Levesque defines a limited version of an ATMS by replacing IMPS with EXPS, which is the set of clauses explicitly believed at $\Re[\text{KB}]$.

Definition 15 (Levesque) *A limited ATMS*

1. $\text{EXPS}[\text{KB}] = \{c \mid c$ *is a tautologous clause or* $\exists c^* \in \text{KB}, c^* \subseteq c\}$.

2. $\mu\text{EXPS}[\text{KB}] = \{c \mid c \in \text{EXPS}[\text{KB}]$ *and for no* $c' \subsetneq c, c' \in \text{EXPS}[\text{KB}]\}$.

3. *Let c be a clause. Then*
$\texttt{lim\_atms}[\text{KB}, c] = \{(q_1 \wedge \ldots \wedge q_k) \mid k \geq 0 \text{ and } (\overline{q_1} \vee \ldots \vee \overline{q_k} \vee c) \in \mu\text{EXPS}[\text{KB}]\}$.[15]

As in the case of implicit belief, Levesque shows that minimal explanations and the result of `lim_atms` coincide if viewed as propositions [Lev89]. `lim_atms`, while considerably weaker than `atms`, has the benefit of being efficiently computable.

Theorem 3 (Levesque) *For any clause c,* $\texttt{lim\_atms}[\text{KB}, c]$ *is computable in* $O(|\text{KB}| \times |c|)$.

Coming back to our original goal of defining a tractable notion of only-knowing-about, we note that `lim_atms` gives us the right characterization of what is known about some atom p (we will look at an arbitrary π in a moment) for a given KB, just as `atms` does in the case of implicit belief.

Theorem 4 *Let p be an atom,* $\Gamma_p = \texttt{lim\_atms}[\text{KB}, p]$*, and* $\Gamma_{\overline{p}} = \texttt{lim\_atms}[\text{KB}, \neg p]$*. Then*

$$\models \mathbf{O}\text{KB} \supset \mathbf{O}\langle p\rangle[\bigwedge_{c \in \Gamma_p} (c \supset p) \wedge \bigwedge_{c \in \Gamma_{\overline{p}}} (c \supset \neg p)].$$

Proof : The proof is a straightforward adaptation of the proof of the analogous result for implicit belief (Theorem 2). ■

More importantly, given the previous two theorems, it is not hard to show that all that is known about p is computable in time linear in the size of KB.

Theorem 5 *All that is known about p with respect to* KB *is computable in* $O(|\text{KB}|)$.

Proof : By Theorem 3, Γ_p and $\Gamma_{\overline{p}}$ can be computed in time $O(|\text{KB}|)$. Also, the size of $\alpha = \bigwedge_{c \in \Gamma_p} (c \supset p) \wedge \bigwedge_{c \in \Gamma_{\overline{p}}} (c \supset \neg p)$ is bounded by the size of KB, that is, α can be constructed in time $O(|\text{KB}|)$. ■

The general case of an arbitrary subject π is not much harder. First, note that Lemma 4.1 also holds in the case of OB^a. Thus, in order to compute what is known about π we only need to compute what is known about each $p \in \pi$ and conjoin the results. This can be done in $O(|\text{KB}| \times |\pi|)$.

Finally, it is perhaps instructive to look at an example that clearly exhibits the reasons why it is so easy to compute what is known about π.

Example 5.1 Let $\text{KB} = \{p, (s \supset q), (p \wedge q) \supset r\}$ and let $\pi = \{r\}$. Then $\mathbf{O}\text{KB} \supset \mathbf{O}\langle\pi\rangle[(p \wedge q) \supset r]$, that is, in order to find out what is known about r we only need to go through the clauses (which are written as implications here) and collect those that mention r. Note that under implicit belief, we would also have to include $(s \supset r)$ since $\mathbf{O}\text{KB} \supset \mathbf{L}(s \supset r)$. In other words, we would have to apply full resolution to find all the relevant clauses.

[15] Levesque considers the more general case of objective sentences in conjunctive normal form as the second argument to `lim_atms`, which he calls `abd`.

6 Summary and future work

In this paper, we proposed a formalization of only-knowing-about, that is, the notion that something is all an agent knows about some subject matter, where *subject matter* was taken to be a set of atoms in a propositional logic. We explored some of the logical properties of this notion and were able to show a tight connection to deKleer's ATMS. Furthermore, by weakening the underlying model of belief, we arrived at a form of only-knowing-about that is efficiently computable.

As for future work, OL^a needs an axiomatization to really understand all of the properties of only-knowing-about. Furthermore, the logic needs to be generalized to the first-order case. This is not at all straightforward, since the propositional case relies heavily on the fact that we can restrict ourselves to clauses rather than arbitrary sentences, an assumption that does not work in the first-order case. Another issue is the multi-agent scenario, where only-knowing-about plays an important role as indicated in the introduction. However, so far there is not even a multi-agent formalization of only-knowing!

Acknowledgements

I would like to thank Wiktor Marek, who suggested the title of the paper as the title of a talk on preliminary results of this work.

References

[AB75] Anderson, A. R. und Belnap, N. D., *Entailment, the logic of relevance and necessity.* Princeton University Press, 1975.

[deK86] de Kleer, J., An assumption-based TMS. *Artificial Intelligence*, **28**(2), 1986, pp. 127–162.

[Dunn 1976] Dunn, J. M., Intuitive Semantics for First-Degree Entailments and Coupled Trees. *Philosophical Studies* **29**, 1976, pp. 149–168.

[HM84] Halpern, J. Y. and Moses, Y. O., Towards a Theory of Knowledge and Ignorance: Preliminary Report, in Proceedings of The Non-Monotonic Workshop, New Paltz, NY, 1984, pp.125–143.

[HM85] Halpern, J. Y. and Moses, Y. O., A Guide to the Modal Logics of Knowledge and Belief. *Proc. of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, CA, 1985, pp. 480–490.

[Hin62] Hintikka, J., *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* Cornell University Press, 1962.

[Kri63] Kripke, S. A., Semantical considerations on modal logic. *Acta Philosophica Fennica* **16**, pp. 83–94, 1963.

[LL88] Lakemeyer, G. and Levesque, H. J., A Tractable Knowledge Representation Service with Full Introspection. *Proc. of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, Asilomar, California, 1988, pp. 145–159.

[Lak92] All I know about Tweety, to appear in: *Proc. of the 2nd International Workshop on Nonmonotonic and Inductive Logics*, Lecture Notes in Computer Science, Springer Verlag, 1992.

[Lev84] Levesque, H. J., A Logic of Implicit and Explicit Belief. *Proc. of the 4th Natinal Conference on Artificial Intelligence (AAAI-84)*, Austin, TX, 1984, pp. 198–202.

[Lev89] Levesque, H. J., A knowledge-level account of abduction. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pp. 1061–1067, Detroit, MA, 1989.

[Lev90] Levesque, H. J., All I Know: A Study in Autoepistemic Logic. *Artificial Intelligence*, North Holland, **42**, 1990, pp. 263–309.

[Moo85] Moore, R., Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence* **25**, 1985, pp. 75–94.

[Poo88] Poole, D., A methodology for using a default and abductive reasoning system. Thechnical Report. Department of Computer Science, University of Waterloo, Waterloo, Ontario, 1988.

[RdK87] Reiter, R. und de Kleer, J., Foundations of assumption-based truth maintenance systems: preliminary report. *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, Seattle, WA, 1987, pp. 183–188.

[SL90] Selman, B. and Leveque H. J., Abductive and default reasoning: a computational core. *Proceedings of the Eigth National Conference on Artificial Intelligence (AAAI-90)*, Boston, MA, 1990, pp. 343–348.