A Logical Account of Relevance

Gerhard Lakemeyer
Institute of Computer Science III
University of Bonn
Romerstr. 164
D-53117 Bonn, Germany
gerhard@cs.uni-bonn.de

Abstract

The study of relevance has gained considerable attention recently in areas as diverse as machine learning and knowledge representation. In this paper, we focus on one particular area, namely relevance in the context of logical theories. We are interested in answering questions like: when is a sentence (or theory) relevant to a set of propositions, or, when is one set of propositions relevant to another given some background theory? The answers are given semantically in terms of a logic of belief and syntactically in terms of prime implicates. Furthermore, rather than merely adding yet another set of definitions of relevance to the many that already exist, we also show close connections to two others that were recently proposed, thus pointing to some common ground at least as far as logical relevance is concerned.

1 Introduction

The study of relevance has gained considerable attention recently in areas as diverse as machine learning and knowledge representation [GS94]. In this paper, we focus on one particular area, namely relevance in the context of propositional logical theories. We are interested in answering questions like: when is a sentence (or theory) relevant to a set of propositions, or, when is one set of propositions relevant to another given some background theory?

In order to get a better feel for what we are getting at, let us consider the following example adapted from [DP94].

a = (rain O **wet)** A (sprinklerjon *D* wet).

Here rain, **wet,** and sprinklerjon stand for the propositions *"it rains", "the ground is wet",* and *"the sprinkler is on",* respectively, *a* clearly seems to tell us something relevant about the proposition rain (as well as sprinklerjon and wet). On the other hand, a seems irrelevant to anything else like Jack_is_happy. When we look a little closer, not everything *ex* conveys is relevant to rain because the sentence contains the extraneous information (sprinklerjon) wet). On the other hand, everything *a* tells us is about the ground being wet and, hence, we want to call *a strictly* relevant to **wet.**

It also makes sense to talk about relevance between propositions relative to a background theory. For example, given *a,* rain should count as relevant to **wet** since the latter is true whenever the former is. On the other hand, there really is no connection between rain and sprinklerjon, since they are either forced to both be false by some other condition (the ground is not wet) or their truth values can vary independently of each other. Therefore, rain and sprinklerjon are not relevant to each other given *a.*

If we view relevance with regard to what a sentence (or theory) tells us about the world, as is done in this paper, then logically equivalent formulations should not change the relevance relation. For example, given a' = (rain D wet) A rain, rain and wet are not considered relevant to each other, since *a'* is logically equivalent to (wet A rain). In this respect, our approach is quite different from work such as [SG87; LFS94], where relevance has a distinct syntactic flavor. There, rain is viewed as relevant to **wet** since rain is used in the syntactic derivation of **wet.**

The formalization of logical relevance chosen in this paper is based on a logic of belief developed earlier [Lak92], which allows us to express that a sentence is believed *about* a set of proposition (also called the *subject matter)* and, more importantly, that a sentence is *all* that is known[1] about a subject matter (only-knowing-about for short). With these concepts we say that a sentence a is relevant to *p* just in case it is impossible to know nothing about *p* assuming *a* is all that is known. Going back to our initial example, *a* is then found to be relevant to rain because knowing only *a* implies that we know

[1]For the purpose of this paper the distinction between knowledge and belief is irrelevant.

something nontrivial about **rain,** namely **(rain $\supset$ wet).** Note the importance of assuming that *only a* is believed and nothing else. For if we merely require *a* to be believed, we do not rule out believing **wet** as well, in which case $\alpha \wedge$ **wet** reduces to **wet** and all relevance to **rain** disappears.

Given a logic of only-knowing-about, it is perhaps not surprising that we are able to express this form of relevance in such a direct manner, since the logic has a built-in primitive notion of relevance in the sense of *aboutness.* It is perhaps more interesting that this primitive notion suffices to provide a reasonable semantics to other varieties of relevance as well. For instance, **rain** is relevant to **wet** relative to *a* because what is known about **rain,** viewed as a sentence, is relevant to **wet.**

Besides defining various forms of relevance sernantically in terms of only-knowing-about, we also provide syntactic characterizations in terms of prime implicates in each case. Furthermore, rather than merely adding yet another set of definitions of relevance to the many that already exist, we also show close connections to two others that were recently proposed [DP94; LR94], thus pointing to some common ground at least as far as logical relevance is concerned.

The rest of the paper is organized as follows. In Section 2, we define the semantics of only-knowing-about, which is a variant of the semantics presented in [Lak92]. Section 3 contains our various definitions of relevance using the logic of the previous section. In Sections 4 and 5, we compare our work to that of Darwiche and Pearl on the one hand and Lin and Reiter on the other. We end the paper with a summary and some concluding remarks.

2 The Logic of Only-Knowing-About

The logic of only-knowing-about, which was originally introduced in [Lak92], extends earlier work by Levesque [Lev90], who formalized what it means to only know a sentence, which can be thought of as the limit-case of only-knowing-about, where the subject matter includes everything the agent has any information about. The semantic framework for all these investigations is possible-world semantics [Kri63; Hin62; HM92].

2.1 From Only-Knowing to Only-Knowing-About

Before introducing the semantics for "all the agent knows about *x* is y," we start out with the simpler case of Levesque's only-knowing. There an agent knows a sentence a, denoted as La, just in case *a* is true in all the worlds which the agent thinks are possible (or accessible). The reader should think of a world simply as a truth assignment for the atomic propositions. (Formal definitions are deferred to Section 2.3 below.) To define only-knowing, Levesque considers another modality N, where **Na** means that a is true in all the impossible (or inaccessible) worlds.[2] While $\mathbf{L}\alpha$ is best understood as "the agent knows *at least* that α is true," Na should be read as "the agent knows *at most* $\neg\alpha$." With that only-knowing α, denoted as $\mathbf{O}\alpha$, reduces to knowing at least α and at most α, that is, $\mathbf{O}\alpha$ holds just in case both $\mathbf{L}\alpha$ and $\mathbf{N}\neg\alpha$ hold.

Let us now consider how to extend these ideas to include a subject matter. Since we confine ourselves to propositional logic, we define a *subject matter* π as a finite set of atomic propositions. For each such π we introduce new modal operators $\mathbf{L}\langle\pi\rangle$, $\mathbf{N}\langle\pi\rangle$, and $\mathbf{O}\langle\pi\rangle$, where $\mathbf{L}\langle\pi\rangle\alpha$ is read as "the agent knows at least α about π," $\mathbf{N}\langle\pi\rangle\alpha$ as "the agent knows at most that α is false about π," and $\mathbf{O}\langle\pi\rangle\alpha$ as "all the agent knows about π is α." As in the case of only-knowing, $\mathbf{O}\langle\pi\rangle\alpha$ can be viewed as shorthand for $\mathbf{L}\langle\pi\rangle\alpha \wedge \mathbf{N}\langle\pi\rangle\neg\alpha$

As for the semantics of these modalities, suppose the beliefs of the agent are given by the set of worlds *M* the agent thinks possible. To find out what the agent knows about π we construct a set of worlds $M|_\pi$ which, intuitively, represents what the agent knows after forgetting everything that is not relevant to π. With that the operators $\mathbf{L}\langle\pi\rangle, \mathbf{N}\langle\pi\rangle$ and $\mathbf{O}\langle\pi\rangle$ are interpreted just like L,N, and O except that we are using $M|_\pi$ instead of *M.* For example, an agent believes a about pat a set of worlds *M* just in case she believes a at $M|_{\{p\}}$

The logic is a slight variant of the one presented in [Lak92j. In particular, we ignore nested modalities, which are not essential for our purposes. The operators $\mathbf{L}\langle\pi\rangle$ and $\mathbf{N}\langle\pi\rangle$ were not used in the previous version of the logic.

2.2 The Language and Other Notation

The primitives of the language are a countably infinite set *V* of atomic propositions (or atoms), the connectives $\vee$, $\neg$, and the modal operators **L,N,0,** $\mathbf{L}\langle\pi\rangle$, $\mathbf{N}\langle\pi\rangle$, and $\mathbf{O}\langle\pi\rangle$ for every every finite set of atomic propositions π with the restriction that none of the modal operators occurs within the scope of another modal operator. Sentences are formed in the usual way from these primitives.[3]

Notation: As usual, literals are either atoms or negated atoms and clauses are disjunctions of literals. We write **false** as an abbreviation for $(p \wedge \neg p)$, where *p* is some atom, and **true** for **-"false.** Given an atomic proposition *p* and and a clause c, we say that *c* mentions *p* just in case either *p* or $\neg p$ occurs in c.

It is often convenient to identify a clause with the set of literals occurring in the clause. A clause *c* is contained in a clause c' $(c \subseteq c')$ if every literal in *c* occurs in *c'.*

[2]Note that here *impossible* does not mean that these worlds are logically impossible. They are merely incompatible with the agent's knowledge.

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

We write $c \subsetneq c'$ instead of $c \subseteq c'$ and $c' \not\subseteq c$. Given a finite set of sentences Γ, $\bigwedge_{\gamma \in \Gamma} \gamma$ denotes the conjunction of all the sentences occurring in Γ. If Γ is empty, $\bigwedge_{\gamma \in \Gamma} \gamma$ denotes **true**. For notational convenience, we sometimes use Γ itself instead of $\bigwedge_{\gamma \in \Gamma} \gamma$ within a sentence.

When referring to singleton subject matters, we usually omit the curly brackets. For example, given an atom p, we write $\mathbf{O}\langle p \rangle$ instead of $\mathbf{O}\langle \{p\} \rangle$ and $M|_p$ instead of $M|_{\{p\}}$.

Finally, a sentence is called *objective* if it contains no modal operator and *subjective* if every atom occurs within the scope of a modal operator.

2.3 A Formal Semantics

Worlds are defined extensionally as propositional truth assignments. Hence for any given set of worlds M (the possible worlds) its complement (the impossible worlds) is always well defined.

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

We begin by reviewing the semantics of Levesque's logic of only-knowing. Let w be a world and M a set of worlds.

$$
\begin{array}{lcl}
M, w \models p & \iff & w(p) = \mathbf{t}, \text{ where } p \text{ is an atom} \\
M, w \models \neg\alpha & \iff & M, w \not\models \alpha \\
M, w \models \alpha \vee \beta & \iff & M, w \models \alpha \text{ or } M, w \models \beta \\
M, w \models \mathbf{L}\alpha & \iff & \text{for all } w' \in M,\ M, w' \models \alpha \\
M, w \models \mathbf{N}\alpha & \iff & \text{for all } w' \notin M,\ M, w' \models \alpha \\
M, w \models \mathbf{O}\alpha & \iff & M, w \models \mathbf{L}\alpha \text{ and } M, w \models \mathbf{N}\neg\alpha
\end{array}
$$

For a given set of worlds M and a subject matter π, we define $M|_\pi$ as the set of all worlds that satisfy precisely the known objective sentences that are about π. Since we are dealing with propositional beliefs only, we can, first of all, confine ourselves to clauses instead of arbitrary objective sentences. The idea to get at only those beliefs that are about π is to consider the smallest clauses which are believed and which mention at least one of the atoms in π. $M|_\pi$ is then simply the set of all worlds that satisfy all of these clauses. Formally:

Definition 2
Let M be a set of worlds and π a subject matter.

1. *A clause c is called M-***minimal** *iff $M \models \mathbf{L}c$ and for all clauses $c' \subsetneq c$, $M \not\models \mathbf{L}c'$.*
2. *A clause c is called M-p-***minimal** *iff c is M-minimal and, in addition, c mentions p.*
3. $\Gamma_{M,\pi} =$
   $\{c \mid c$ *is an M-p-minimal clause for some* $p \in \pi.\}$
4. $M|_\pi = \{w \mid w \models c$ *for all* $c \in \Gamma_{M,\pi}\}$.

By restricting ourselves to M-p-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.

Given these definitions of what it means to forget irrelevant things, we obtain the following semantic rules for knowing and only-knowing about.

$$
\begin{array}{lcl}
M, w \models \mathbf{L}\langle\pi\rangle\alpha & \iff & M|_\pi, w \models \mathbf{L}\alpha \\
M, w \models \mathbf{N}\langle\pi\rangle\alpha & \iff & M|_\pi, w \models \mathbf{N}\alpha \\
M, w \models \mathbf{O}\langle\pi\rangle\alpha & \iff & M, w \models \mathbf{L}\langle\pi\rangle\alpha \wedge \mathbf{N}\langle\pi\rangle\neg\alpha
\end{array}
$$

Note: In the original definition of $\mathbf{O}\langle\pi\rangle$ in [Lak92], $M, w \models \mathbf{O}\langle\pi\rangle\alpha$ was defined as $M|_\pi, w \models \mathbf{O}\alpha$ and $M, w \models \mathbf{L}\alpha$. The restriction "$M, w \models \mathbf{L}\alpha$" was necessary to prevent unintuitive properties in the case of nested beliefs. In the unnested case, as in this paper, these problems do not occur and we get by with the simpler definition. In the following we often write $w \models \alpha$ instead of $M, w \models \alpha$ if α is objective and $M \models \alpha$ if α is subjective.

Logical implication, validity, and satisfiability are defined as usual: A set of sentences Γ *logically implies* a sentence α ($\Gamma \models \alpha$) iff for all worlds w and for all 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 $\{\} \models \alpha$. α is satisfiable iff $\neg\alpha$ is not valid.[4]

2.4 Prime implicates

Given all an agent knows, her knowledge about a subject matter can be characterized succinctly using *prime implicates*. Given a sentence α (the agent's knowledge), the prime implicates of α are simply the smallest clauses (in terms of set inclusion) that are logically implied by α. Formally:

Definition 3 (Prime Implicates)
Let α be an objective sentence. A clause c is called a prime implicate of α iff

1. $\models \alpha \supset c$ *and*
2. *for all* $c' \subsetneq c$, $\not\models \alpha \supset c'$.

For any atom p, let $\mathcal{P}(\alpha, p) = \{c \mid c$ is a prime implicate of α mentioning $p\}$ and $\mathcal{P}(\alpha) = \bigcup_{p \in \alpha} \mathcal{P}(\alpha, p)$.[5]

For example, if $\alpha = (p \vee q) \wedge (p \vee r \vee s) \wedge \neg s$, then $\mathcal{P}(\alpha, p) = \{(p \vee q), (p \vee r), (p \vee \neg p)\}$. In cases like $\alpha = q$, where p is not contained in α at all, $\mathcal{P}(\alpha, p) = \{(p \vee \neg p)\}$.

It is easy to see that for any given α and p, $\mathcal{P}(\alpha, p)$ and $\mathcal{P}(\alpha)$ are finite assuming we identify clauses with sets of literals and hence eliminate redundancies. What

[4] In his original definition of only-knowing [Lev90], Levesque considers only what he calls *maximal* sets of worlds instead of arbitrary ones. Here we simply ignore this issue since it is not important for the purpose of this paper.

[5] $p \in \alpha$ stands for p occurs in α.

is known about π relative to a sentence (background theory) α has a simple characterization in terms of prime implicates of α, as the following theorem shows.[6]

Lemma 1 *Let $M \models \mathbf{O}\alpha$. Then for any atom p, $c \in \mathcal{P}(\alpha,p)$ iff c is an M-p-minimal clause.*

Theorem 2 $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi\rangle\beta$ *iff* $\models \beta \equiv \bigwedge_{p\in\pi} \bigwedge_{\gamma\in\mathcal{P}(\alpha,p)} \gamma$.

Proof: Let $M \models \mathbf{O}\alpha$ and let $\beta^* = \bigwedge_{p\in\pi} \bigwedge_{\gamma\in\mathcal{P}(\alpha,p)} \gamma$. We show that $M \models \mathbf{O}\langle\pi\rangle\beta^*$. Together with the fact that a sentence that is only-believed about π is unique up to equivalence, the theorem follows.

To show that $M \models \mathbf{O}\langle\pi\rangle\beta^*$, consider $M|_\pi = \{w \mid w \models c \text{ for all } c \in \Gamma_{M,\pi}\}$ with $\Gamma_{M,\pi} = \{c \mid c \text{ is an } M\text{-}p\text{-minimal clause for some } p \in \pi\}$. By Lemma 1, $\Gamma_{M,\pi} = \{c \mid c \in \mathcal{P}(\alpha,p) \text{ for some } p \in \pi\}$. Therefore, $M|_\pi \models \mathbf{O}\beta^*$ and thus $M \models \mathbf{O}\langle\pi\rangle\beta^*$. ∎

3 Varieties of Relevance

In this section, we will see how logical relevance in the sense we introduced in Section 1 can be captured in a natural way using only-knowing-about.[7] We begin by defining what it means for a sentence to be relevant to some subject matter. The intuition behind α being relevant to π is that α must contain nontrivial information about π. Our logic allows us to express this directly.

Definition 4 *An objective sentence α is relevant to a subject matter π iff $\models \mathbf{O}\alpha \supset \neg\mathbf{O}\langle\pi\rangle(p \vee \neg p)$.*

Example 1 *While $\neg p$ and $(p \supset q)\wedge(q \supset r)$ are relevant to p, $(q \supset r)$ and $p \supset (q \supset p)$ are not.*

Theorem 3 *α is relevant to π iff there is some $\gamma \in \mathcal{P}(\alpha)$ such that $\not\models\gamma$ and γ mentions some $p \in \pi$.*

Proof: The theorem follows immediately from Theorem 2. ∎

While the previous definition, in a sense, only requires part of the sentence to be about π, we can be even more restrictive and require that everything α tells us is about π in a relevant way and hence arrive at the notion of strict relevance.[8]

Definition 5 *An objective sentence α is strictly relevant to a subject matter π iff $\not\models\alpha$ and $\mathbf{O}\langle\pi\rangle\alpha$ is satisfiable.*

[6] While the result follows from Theorem 2 in [Lak92], where it is shown how to compute what is only-known about π using deKleer's ATMS[deK86; RdK87], we give a much simpler direct proof here.

[7] Some of the material of this section appeared in [Lak94].

[8] This definition was first introduced in [Lak93].

Example 2 *$(p \supset q)\wedge(q \supset r)$ is not strictly relevant to p because $(q \supset r)$ is not about p. However, $(p \equiv q)\wedge(q \supset r)$ is strictly relevant to p. This time $(q \supset r)$ is recognized as being about p since p and q are assumed to be equivalent.*

Theorem 4 *Let α be an objective sentence such that $\not\models\alpha$. Then the following statements are equivalent.*

1. *α is strictly relevant to π.*
2. $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi\rangle\alpha$.
3. $\models \alpha \equiv \bigwedge_{p\in\pi} \bigwedge_{\gamma\in\mathcal{P}(\alpha,p)} \gamma$.

Proof: The equivalence of (2) and (3) follows immediately from Theorem 2. We now show that (1) iff (2). If direction: By assumption, $\not\models\alpha$. Also, $\mathbf{O}\alpha$ is satisfiable for any objective α. Since $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi\rangle\alpha$ by assumption, $\mathbf{O}\langle\pi\rangle\alpha$ is satisfiable as well.

Only-if direction: Let $\mathbf{O}\langle\pi\rangle\alpha$ be satisfiable. Hence there is an M such that $M|_\pi \models \mathbf{O}\alpha$, that is, $M|_\pi = \{w \mid w \models \alpha\}$. To show that $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi\rangle\alpha$, let $M^* \models \mathbf{O}\alpha$. Hence $M^* = M|_\pi$. One can easily show that $M|_\pi = (M|_\pi)|_\pi$ for any M. Thus $M^* = M^*|_\pi$ and, therefore, $M^*|_\pi \models \mathbf{O}\alpha$, which immediately implies $M^* \models \mathbf{O}\langle\pi\rangle\alpha$. ∎

Next we want to express that a subject matter is relevant to another relative to some background theory.

Definition 6 *Let π_1 and π_2 be sets of atoms and α and β objective sentences such that $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi_2\rangle\beta$. π_1 is relevant to π_2 with respect to α [$R_\alpha(\pi_1,\pi_2)$] iff $\pi_1 \cap \pi_2 \neq \{\}$ or β is relevant to π_1.*

In other words, π_1 is relevant to π_2 if whatever is known about π_2 contains some nontrivial information about π_1.

Example 3
Let $\alpha = (p \vee q) \wedge (q \vee r)$. Since $\models \mathbf{O}\alpha \supset \mathbf{O}\langle q\rangle\alpha$ and $\models \mathbf{O}\alpha \supset \mathbf{O}\langle p\rangle(p \vee q)$, we obtain immediately that p is relevant to q. Similarly, q is relevant to r. However, p is not relevant to r, since $\models \mathbf{O}\alpha \supset \mathbf{O}\langle r\rangle(q \vee r)$ and $\models \mathbf{O}(q \vee r) \supset \mathbf{O}\langle p\rangle(p \vee \neg p)$.

Theorem 5 *Let π_1 and π_2 be disjoint sets of atoms. $R_\alpha(\pi_1,\pi_2)$ iff there is a $\gamma \in \mathcal{P}(\alpha)$ such that γ mentions atoms from both π_1 and π_2.*

Proof: Let β be such that $\models \mathbf{O}\alpha \supset \mathbf{O}\langle\pi_2\rangle\beta$.

If direction: Let $\gamma \in \mathcal{P}(\alpha)$ such that γ mentions both π_1 and π_2. Then γ is also a prime implicate of β. Since γ mentions atoms from both π_1 and π_2, γ is not a tautology. Thus, by Theorem 3, β is relevant to π_1.

Only-if direction: Let $R_\alpha(\pi_1,\pi_2)$. Then β is relevant to π_1 and by Theorem 3, there is a $c \in \mathcal{P}(\beta)$ such that c mentions some atom in π_1. Also $\models \beta \equiv \bigwedge_{p\in\pi_2} \bigwedge_{\gamma\in\mathcal{P}(\alpha,p)} \gamma$. Assume none of these γ mentions π_1. Then none of the prime implicates of β mentions atoms from π_1, a contradiction. ∎

The relation $R_\alpha(\pi_1,\pi_2)$ is obviously reflexive by definition. While symmetry is not obvious from the definition, it nevertheless follows immediately from Theorem 5, that is, if π_1 is relevant to π_2 with respect to α, then π_2 is relevant to π_1. Note, however, that transitivity does not hold. Example 3 provides a counterexample.

While the previous definition requires π_1 and π_2 only to be weakly connected to each other, the following and last definition forces this connection to be much stronger. In particular, we require that whatever is known about π_1 is also known about π_2.

Definition 7
π_1 is subsumed by π_2 with respect to α ($\pi_1 \prec_\alpha \pi_2$) iff $\models \mathbf{O}\alpha \supset (\mathbf{L}(\pi_1)\beta \supset \mathbf{L}(\pi_2)\beta)$ for all objective β. π_1 and π_2 are equivalent with respect to α ($\pi_1 \approx_\alpha \pi_2$) iff $\pi_1 \prec_\alpha \pi_2$ and $\pi_2 \prec_\alpha \pi_1$.

It is easy to see that $\models \alpha \supset (p \equiv q)$ implies $p \approx_\alpha q$. Note, however, that the converse does not hold. For example, $p \approx_\alpha q$ holds even for $\alpha = (p \supset q)$. With Theorem 2 we obtain the following

Theorem 6 *$\pi_1 \prec_\alpha \pi_2$ iff $\mathcal{P}(\alpha, \pi_1) \subseteq \mathcal{P}(\alpha, \pi_2)$.*

As an aside, while the syntactic characterizations of our definitions of relevance do not appeal to modal logic at all, which may be a consolation for those wary of non-classical logic, the modal characterization has its advantages. For one, it provides a model theoretic account of relevance. For another, characterizing relevance as certain valid sentences opens the door to reason about relevance ***within*** the logic itself. (See also [Lak95] for a proof theoretic characterization of relevance using only-knowing-about.)

Besides the four definitions of relevance given above others are plausible as well. For example, we could replace *relevant* by *strictly relevant* in Definition 6. However, we won't pursue this issue further at this point. Instead we take a closer look at how our definitions, in particular, Definition 6, relate to other work on relevance in the literature.

4 Conditional independence

Recently Darwiche and Pearl (DP) defined *conditional independence* for disjoint sets of atomic propositions [DP94] in the context of a logical theory. This notion was proposed in close analogy to conditional independence in the probabilistic literature, where it is recognized as being of central importance, in particular in areas such as Bayesian Networks [Pea88]. Darwiche [Dar94] has pointed out how (logical) conditional independence gives rise to a natural notion of relevance, or better, irrelevance.

In this section we will show a close connection between DP's notion of relevance and ours, in particular, Definition 6, where we define relevance between two sets of atoms relative to a given background theory.

Notation: Following DP, we use capital letters X, Y, and Z to denote (disjoint) sets of atoms. An instantiation $\hat{X}$ of Xdenotes a consistent set of literals over the atoms X, that is, $\hat{X}$ contains p or $\neg p$ for every $p \in X$. For any instantiation $\hat{X}$ of X, $\overline{\hat{X}}$ denotes the set of literals which are complements of literals in $\hat{X}$. For example. if $X = \{p, q, r\}$ and $\hat{X} = \{\neg p, q, \neg r\}$, then $\overline{\hat{X}} = \{p, \neg q, r\}$. Given sets of literals $\hat{X}$ and $\hat{Y}$, we write $\{\hat{X}, \hat{Y}\}$ instead of $\hat{X} \cup \hat{Y}$. Also, depending on the context, a set of literals denotes a disjunction of literals or a conjunction. As a rule, whenever the set is added to a theory, it is understood conjunctively. Whenever it appears on the right-hand-side of $\models$, it is understood disjunctively. Finally, all sentences and sets of sentences are objective unless noted otherwise.

Definition 8 (Conditional Independence)
Let X, Y, and Z be disjoint sets of atomic propositions and Δ a finite set of sentences. Δ finds X independent of Y given Z [$I_\Delta(X, Z, Y)$] iff for every instantiation $\hat{X}$, $\hat{Y}$, and $\hat{Z}$of X, Y, and Z, respectively, the logical consistency of $\Delta \cup \{\hat{X}, \hat{Z}\}$ and $\Delta \cup \{\hat{Y}, \hat{Z}\}$ implies the logical consistency of $\Delta \cup \{\hat{X}, \hat{Y}, \hat{Z}\}$.

Intuitively, X is independent of Y given Z wrt. Δ if, given full information about Z (namely $\hat{Z}$), adding information about X to Δ in no way effects the information about Y.[9] Using one of DP's examples,

$$\Delta = \{(\mathtt{rain} \vee \mathtt{sprinkler\_on}) \equiv \mathtt{wet}\}$$

finds $\{\mathtt{rain}\}$ independent of $\{\mathtt{sprinkler\_on}\}$, but finds them dependent given $\{\mathtt{wet}\}$.

While our notion of relevance between sets of atoms is only a two-place relation, there is nevertheless a tight connection with conditional independence. This is best seen using our example from Section 1 rephrased as $\Delta = \{(\mathtt{rain} \supset \mathtt{wet}), (\mathtt{sprinkler\_on} \supset \mathtt{wet})\}$. Clearly, `rain` and `sprinkler_on` are not relevant to each other, that is, $R_\Delta(\{\mathtt{rain}\}, \{\mathtt{sprinkler\_on}\})$ does not hold. On the other hand, we obtain that for all possible Z, namely $\{\}$ and $\{\mathtt{wet}\}$, $I_\Delta(\{\mathtt{rain}\}, Z, \{\mathtt{sprinkler\_on}\})$ holds. Intuitively, this says that not being relevant to one another is the strongest case of conditional independence, where one quantifies over all possible Z. It turns out that this property holds in general.

Theorem 7 *Let X and Y be disjoint sets of atoms and Δ a finite set of sentences. Then $R_\Delta(X, Y)$ iff there exists a Z such that $\neg I_\Delta(X, Z, Y)$.*[10]

Proof : (Here we only prove the only-if direction.) Thus, if we assume $R_\Delta(X, Y)$, then, by Theorem 5, there is a $c \in \mathcal{P}(\Delta)$ such that c mentions atoms from both X and Y. Thus $c = \{\hat{Z}, \hat{X}_1, \hat{Y}_1\}$ for some X_1, Y_1, and Z such that $X_1 \subseteq X, Y_1 \subseteq Y, X_1 \neq \{\}, and Y_1 \neq \{\}$. We will now show that $\neg I_\Delta(X, Z, Y)$.

[9]This definition is also used in the relational database literature to characterize what they call *embedded multivalued dependencies*.

[10]$\neg I_\Delta(X, Z, Y)$ is short for $I_\Delta(X, Z, Y)$ *does not hold*.

Since $c \in \mathcal{P}(\Delta)$, $\Delta \not\models \{\hat{Z}, \hat{X_1}\}$ and $\Delta \not\models \{\hat{Z}, \hat{Y_1}\}$, but $\Delta \models \{\hat{Z}, \hat{X}_1, \hat{Y}_1\}$. If we choose $\overline{\hat{Z}}$, $\overline{\hat{X}_1}$, and $\overline{\hat{Y}_1}$ as instantiations of Z, X_1, and Y_1, respectively, we obtain that $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{X}_1}\}$ and $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{Y}_1}\}$ are both consistent, while $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{X}_1}, \overline{\hat{Y}_1}\}$ is inconsistent. Hence $\neg I_\Delta(X_1, Z, Y_1)$.

Since $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{X}_1}\}$ and $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{Y}_1}\}$ are consistent, there are $\hat{X}$ and $\hat{Y}$ with $\overline{\hat{X}_1} \subseteq \hat{X}$ and $\overline{\hat{Y}_1} \subseteq \hat{Y}$ such that $\Delta \cup \{\overline{\hat{Z}}, \hat{X}\}$ and $\Delta \cup \{\overline{\hat{Z}}, \hat{Y}\}$ are consistent. Since $\Delta \cup \{\overline{\hat{Z}}, \overline{\hat{X}_1}, \overline{\hat{Y}_1}\}$ is inconsistent, then so is $\Delta \cup \{\overline{\hat{Z}}, \hat{X}, \hat{Y}\}$. Hence $\neg I_\Delta(X, Z, Y)$. ■

5 Relevance according to Lin and Reiter

In [LR94], Lin and Reiter (LR) introduce a form of forgetting all the information about a proposition in a given set of sentences, which they use, among other things, to define notions of *remembering* and *(ir-)relevance*.[11]

Our first task will be to show that LR's remembering is related, but not identical to only-knowing-about. The difference arises because LR define remembering what is known about an atomic proposition p in terms of forgetting *everything* about propositions other than p, which may have the side effect of erasing information about p as well. For example, if $\Delta = \{(p \vee q)\}$, then nothing is remembered about p since forgetting everything about q in the LR sense removes the clause $(p \vee q)$. On the other hand, according to the logic of only-knowing-about, if all you know is Δ, then all you know about p is still Δ. Despite these differences, Theorem 9 below characterizes a special case where only-knowing-about and remembering coincide. We begin by formally introducing LR's notion of forgetting.

Definition 9 (Forgetting)
Let Δ be a set of sentences and p an atomic proposition. Let Δ_p^+ be the result of replacing every occurrence of p in Δ by **true** *and let Δ_p^- be the result of replacing every occurrence of p in Δ by false.*
Then $\mathbf{forget}(\Delta; p) = (\Delta_p^+ \vee \Delta_p^-)$.[12]

Example 4 *If* $\Delta = \{(p \vee q), (\neg p \vee r), s\}$, *then* $\mathbf{forget}(\Delta; p) = \{(q \vee r), s\}$.

The following lemma should help to get a better feel for the LR notion of forgetting.

Lemma 8 $\mathbf{forget}(\Delta; p)$ *is logically equivalent to* $\{c \in \mathcal{P}(\Delta) \mid c$ *does not mention* $p\}$.

[11] While LR treat first-order formulas, we restrict ourselves to the propositional case.

[12] LR first introduce a semantic definition of $\mathbf{forget}(\Delta; p)$ and then show that this syntactic characterization coincides with the semantic one. For simplicity, we ignore the semantic part.

For atoms $p_1, \ldots, p_k$, LR define $\mathbf{forget}(\Delta; p_1, \ldots, p_k)$ as $\mathbf{forget}(\mathbf{forget}(\Delta; p_1, \ldots, p_{k-1}); p_k)$. (They show that the order of the atoms does not matter.)

Given **forget**, LR define $\mathbf{remember}(\Delta; p_1, \ldots, p_k)$ as a function, which forgets all information about atoms other than $p_1, \ldots, p_k$.

Definition 10 (Remembering)
Let Δ be a set of sentences such that the only atoms occurring in Δ are $p_1, \ldots, p_k$ and $p_{k+1}, \ldots, p_n$. Then $\mathbf{remember}(\Delta; p_1, \ldots, p_k) = \mathbf{forget}(\Delta; p_{k+1}, \ldots, p_n)$.

While **remember** looks a lot like only-knowing-about, there are important differences as the following example demonstrates.

Example 5
Let us consider the same Δ as in Example 4. It is easy to see that $\models \mathbf{O}(\Delta) \supset \mathbf{O}\langle p\rangle[(p \vee q) \wedge (\neg p \vee r)]$.*! On the other hand,* $\mathbf{remember}(\Delta; p) = \mathbf{true}$*!*

The problem is that $\mathbf{remember}(\Delta; p)$, when forgetting about q and r, also removes the clauses $(p \vee q)$ and $(\neg p \vee r)$. The $\mathbf{O}\langle p\rangle$-operator, on the other hand, is more sensitive, and hangs on to those clauses since they tell us something about p.

However, if we add q and r to the subject matter, both $\mathbf{remember}(\Delta; p, q, r)$ and $\mathbf{O}\langle p, q, r\rangle$ produce the same result.

The last remark can be generalized as follows.

Theorem 9 *Let Δ be a set of sentences and $\pi = \{p_1, \ldots, p_k\}$. If every prime implicate of Δ that mentions some $p \in \pi$ mentions* only *atoms in π, then*

$$\models \mathbf{O}(\Delta) \supset \mathbf{O}\langle \pi\rangle(\mathbf{remember}(\Delta; p_1, \ldots, p_k)).$$

Based on **forget** LR go on to define a notion of (ir-)relevance relative to answering a query posed to a knowledge base and the sentences that could possibly be learned (= added to the knowledge base) in the future. Ignoring future information for a moment, an atom p is said to be irrelevant for answering q in the knowledge base Δ just in case Δ and $\mathbf{forget}(\Delta; p)$ produce the same answer for q. Consider the following example, where $\Delta = \{(p \supset q), r\}$ and q is the query. Intuitively, while r seems clearly irrelevant to answering q, p should somehow count as relevant. However, both Δ and $\mathbf{forget}(\Delta; p) = \{r\}$ produce the same answer when asking whether q holds, namely *unknown*. This is where possible future extensions of Δ come in. For example, if p is among the sentences that could be added at some point, then $\Delta \cup \{p\}$ and $\mathbf{forget}(\Delta; p) \cup \{p\}$ differ with respect to q, since the former would produce the answer *yes* and the latter *unknown*. On the other hand, Δ and $\mathbf{forget}(\Delta; r)$ would always produce the same answers for q no matter what information is added to them. Hence by considering future extensions of a knowledge base, LR obtain a reasonable notion of distinguishing relevant from irrelevant information when answering a query.

After defining the LR notion of relevance formally, We will show how $R_\Delta(\pi_1, \pi_2)$ can be expressed within this framework.

Definition 11 *Let Δ_1 and Δ_2 be two finite sets of sentences, $\mathcal{L}$ a set of learnable sentences and α a sentence.*

1. *Δ_1 and Δ_2 are equivalent wrt. α ($\Delta_1 \approx_\alpha \Delta_2$) iff*
   (a) $\Delta_1 \models \alpha$ iff $\Delta_2 \models \alpha$ and
   (b) $\Delta_1 \models \neg\alpha$ iff $\Delta_2 \models \neg\alpha$.
2. *Δ_1 and Δ_2 are equivalent wrt. α and $\mathcal{L}$ ($\Delta_1 \approx_{\alpha,\mathcal{L}} \Delta_2$) iff for all $\mathcal{L}' \subseteq \mathcal{L}$, $\Delta_1 \cup \mathcal{L}' \approx_\alpha \Delta_2 \cup \mathcal{L}'$ whenever $\Delta_1 \cup \mathcal{L}'$ and $\Delta_2 \cup \mathcal{L}'$ are consistent.*

Definition 12 *Let Δ be a finite set of sentences, $\mathcal{L}$ a set of learnable sentences (not necessarily finite), p an atomic proposition, and α a sentence.*

Then p is irrelevant for answering α in Δ iff $\Delta \approx_{\alpha,\mathcal{L}} \mathtt{forget}(\Delta; p)$.

Theorem 10 *Let Δ be a set of sentences, π_1 and π_2 disjoint sets of atoms, and $\mathcal{L}$ the set of all literals. $R_\Delta(\pi_1, \pi_2)$ does not hold iff for all sentences α mentioning only atoms in π_2 and for all $p \in \pi_1$, p is irrelevant for answering α in Δ.*

6 Conclusion

In this paper we proposed various definitions of relevance within a logic of only-knowing-about. We showed that the semantic definitions have simple characterizations in terms of prime implicates. Furthermore, we demonstrated tight connections between our work and that of Darwiche and Pearl on the one hand and Lin and Reiter on the other. While all three approaches evolved independently with different motivations in mind, this paper suggests that there is common ground among them. However, more work is needed to map out their exact relationships.

In future work, our investigations of logical relevance should be extended to the first-order case, which would require a first-order version of the logic of only-knowing-about. This would allow an even better comparison with LR's work, which is already first-order. Judging from LR's experience, though, there may well be nasty surprises along the way. For example, the result of forgetting a predicate the LR way is not always first-order representable.

Acknowledgements

I would like to thank Adrian Darwiche, Adam Grove, Fangzhen Lin, and Ray Reiter for discussions, which helped clarify many ideas presented in this paper. 1 am also grateful to the anonymous referees, whose suggestions helped to improve the presentation of the paper.

References

[Dar94] Darwiche, A., A logical notion of conditional independence: properties and applications, in *Proc. of the AAAI Fall Symposium on Relevance,* New Orleans, 1994, pp. 36-40.

[DP94] Darwiche, A. **and** Pearl, J., Symbolic **Causal** Networks, in Proc. AAAI-94, AAAI-Press, 1994, pp. 238-244.

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

[GS94] Greiner, R. and Subramaniam, D. (Eds.), Proceedings of the AAAI Fall Symposium on Relevance, New Orleans, 1994.

[HM92] Halpern, J. Y. and Moses, Y. O., A Guide to completeness and complexity for modal logics of knowledge and belief, *Artificial Intelligence,* 54, pp. 319-379, 1992.

[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, 1963, pp. 83-94.

[Lak92] Lakemeyer, G., All You ever Wanted to Know about Tweety, in *Proc. of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92),* Morgan Kaufmann, San Mateo, CA, 1992, pp. 639-648.

[Lak93] Lakemeyer, G., All They Know About, *Proc. of the 11th National Conference on Artificial Intelligence (AAAI-93),* MIT Press, Cambridge, MA, 1993, pp. 662-667.

[Lak94] Lakemeyer, G., Relevance in a Logic of Only Knowing About and its Axiomatization, in *Proc. of the AAAI Fall Symposium on Relevance,* New Orleans, 1994, pp. 149-153.

[Lak95] Lakemeyer, G., Relevance in a Logic of Only-Knowing-About and its Axiomatization, to appear in: A. Laux and H. Wansing (eds.) *Knowledge and Belief in Philosophy and Artificial Intelligence,* Akademie-Verlag, Berlin, 1995.

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

[LFS94] Levy, A., Fikes, R., and Sagiv, Y., A Proof-Theoretic Approach to Irrelevance: Foundations and Applications, in *Proc. of the AAAI Fall Symposium on Relevance,* New Orleans, 1994, pp. 149-153.

[LR94] Lin, Fangzhen and Reiter, R., Forget It!, in *Proc. of the AAAI Fall Symposium on Relevance,* New Orleans, 1994, pp. 154-159.

[Pea88] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Francisco, 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.

[SG87] Subramaniam, D. and Genesereth, M. R., The Relevance of Irrelevance, in *Proc. of the 10th Int. Joint Conf. on Artificial Intelligence (IJCAI-87),* Morgan Kaufmann, 1987, pp. 416-422.