A Note on Epistemology and Logical Artificial Intelligence

Gregory R. Wheeler and Luís Moniz Pereira
Centro de Inteligência Artificial (CENTRIA)
Departamento de Informática, Universidade Nova de Lisboa
2829-516 Caparica, Portugal
{greg,lmp}@di.fct.unl.pt

August 18, 2003

Abstract

Contemporary analytical epistemology and logical artificial intelligence are complementary disciplines. This is most apparent when considering the structure of what we call *epistemic* relations, among which inferential are of special concern. In this essay we discuss key correspondence results between semantics for logic programs and default logic. Our aim is to show that there now exist sufficient theoretical foundations within logical artificial intelligence to warrant use of non-monotonic logics as analytical tools within the theory of knowledge.

1 Introduction

Traditional epistemology occupies itself primarily with two sorts of problems. The first concerns the analysis of fundamental epistemic notions, such as *justification*, *evidence* and perhaps also *belief*, along with the analysis of key epistemic relations that appear to involve these concepts, like *is warranted by*, *supports*, and *is reasonable to infer*. In assembling these accounts into a theory, the aim of this project is to give an analysis of *knowledge*—what it is to know a proposition, like when each of us says 'I know I have two hands'.[1]

The other chief concern is the challenge posed by skeptical arguments to the possibility of having knowledge. While there are varieties of philosophical skepticism, a historically significant version concerns the possibility of empirical knowledge about the external world, such as our respective claims of knowing to have two hands. Knowledge claims such as these are justified by our experience, yet it is conceivable that we haven't hands at all. Perhaps instead we each are a brain in a vat, electrochemically deceived into believing in his two-handedness.

[1]For a brief overview of the current state of traditional epistemology, see Jim Pryor's [Prior 2001], which also contains an excellent bibliography.

The serious problem raised by the problem of skepticism is whether in giving an account of knowledge that is refined enough to distinguish wholesale deception from genuine knowledge claims we in fact filter out entire classes of claims from ever being classified as knowledge, such as empirical claims about the world.[2]

While there remain disputes, both over proposals that analyze knowledge on the one hand and various strategies for refuting skepticism on the other, one broad consensus seems to hold among contemporary epistemologists: almost everyone agrees that *Cartesian foundationalism* is not a viable option. Cartesian foundationalism is a particular version of foundationalism, one that holds that knowledge of one's two handedness, say, is derived from basic statements about his own sensations, of which knowledge is supposed indubitable. However, no one thinks that sensations provide infallible reports from the external world since no formulation of the basic sense-statement idea seems to escape skeptical challenge. More importantly, it is no longer believed that epistemic notions behave like truth does in valid derivation—a position that has significant ramifications for the study of epistemic relations, particularly inference relations. Justification is conferrable by induction, which is necessarily not truth preserving. Furthermore, justification is not necessarily conferred to the logical consequences of our beliefs nor does it, when conferred to a true belief by derivation, necessarily guarantee knowledge of that derived belief.

The dimensions of this last point—how fundamentally different justification propagation is from truth preservation—did not begin to become apparent until the 1960's. It was during this decade that several epistemic paradoxes were articulated, including the paradox of the knower [Kaplan and Montague 1960; Cross 2000; Uzquiano, forthcoming] and the paradoxes of rational acceptance, namely the lottery [Kyburg 1961, 1997] and the preface [Makinson 1965; Pollock 1986; Conee 1987]. Each paradox shows that very plausible minimal conditions—on the behavior of a knowledge predicate and those thought necessary for rational acceptance—lead to contradiction. While it is still disputed which conditions should be dropped to resolve each paradox, the lesson we draw from these paradoxes is that closure operations on languages modeling epistemic notions are not isomorphic to any closure operations of classical first-order logic.

Conceptual studies such as Edmund Gettier's famously short "Is Justified True Belief Knowledge?" [Gettier 1963] suggest another reason for thinking that epistemic notions are propagated unlike truth under logical consequence. Gettier's essay brought attention to cases where a justified but false belief may confer justification, by simple derivation, to statements believed but true by chance.[3] So in addition to the problem of skepticism, Gettier cases present

[2]For a recent collection of papers on skepticism, see [DeRose and Warfield, 1999].

[3]One of Gettier's two counter-examples runs as follows. Suppose Smith has very strong evidence for the proposition A, *Jones owns a Ford*. Smith's evidence might include that Jones has always owned a car in the past, it has always been a Ford, and that Smith has just accepted an offer of a ride from Jones who is driving a Ford. We are then asked to imagine another friend of Smith, Brown, whose whereabouts are completely unknown to Smith. Smith selects a place at random and entertains the following proposition B, *Jones owns a Ford or Brown is in Barcelona.* Since A entails B and let us suppose that Smith grasps this entailment, Smith is justified to believe B. But now imagine that in fact Jones does *not* own a Ford; the present

another obstacle within epistemology—one that affects theories of *justification*. Since Gettier, the trick has been to formulate a theory of justification that is strict enough to avoid counting Gettier-style counter examples as cases of knowledge while at once flexible enough to ensure correct classification of common empirical knowledge claims as being justified.

Now, epistemologists are right to stress the differences between logical consequence and inference. All of us are creatures adept at drawing defeasible inferences from information introduced to us by our senses: it is a restricted case when we deduce a conclusion from an explicitly held belief whose contents are an experience.[4] But it is a mistake to think logic plays no role in modeling inference relations [cf. Harman 2001]. Even though there are notable exceptions, most current philosophical theories of knowledge are advanced as though logic offered little analytical insight into the structure of the relations mentioned in each theory, including inference relations. It is standard methodological practice for philosophers to offer theories of justification assembled from a conceptual analysis of both epistemic concepts and epistemic relations, where the behavior of epistemic relations—including inference relations—is described rather than formally defined.

That contemporary epistemology has neglected logic as an analytical research tool may be illustrated by considering the main methodological dispute to exercise the field over the last three decades. Since the publication of W. V. O. Quine's "Epistemology Naturalized" [Quine 1969], epistemologists have been arguing whether their proper home is in psychology departments rather than philosophy departments. Methodological naturalism (in epistemology) is

---

car he is driving is a rented car. Furthermore, by chance, suppose Brown is in Barcelona. So, B is true. However, it no longer appears that Smith knows B. It should be noted that Gettier-style counter examples do not depend upon the justification-conferring belief being false [Feldman 1974].

[4]More on defeasible inference to follow. And we acknowledge the psychological ability we all share to draw reasonable inferences without the slightest awareness of the explicit grounds we have for doing so. However, there are cases where we do evaluate the explicit grounds available for drawing an inference, namely when we consider arguments. Our focus is this class of restricted cases. Finally, the conceptual distinction between beliefs and their contents may be illustrated by considering the difference between having a headache and believing that one's head aches. The content of the belief that one's head aches is having a headache. Notice that having a headache is good grounds for believing one's head aches, but that it is peculiar to cite the belief that one's head aches for grounds to infer that one has a headache. An epistemic relation (and perhaps also a causal relation) holds between (from) a non-propositional experience, a pain in the head, and (to) a doxastic state, a belief that one's head hurts, whose content is the experience of pain in the head.

That non-propositional items may stand in epistemic relations to beliefs we may have is a non-trivial point for knowledge representation. In some dynamic circumstances we appear to draw inferences from graphical or geometric representations of information much better than when that information is represented in propositional form. Meteorologists reach conclusions from weather maps that they are unable to draw from an array of meteorological data represented in propositional form [Hoffman 1991] and air traffic controllers at the busiest airports still rely upon slips of paper, each denoting an aircraft and moved around a controller's field of vision to represent traffic in his sector, from which he may draw inferences about the flow of traffic, degree or distribution of congestion, and ranking of conflicts to resolve [Sellen and Harper 2001].

the view that the results and methods of the cognitive sciences are relevant to doing traditional epistemology. What is interesting about this dispute for our purposes is to notice the relatively narrow scope of the disagreement between 'naturalists' and 'anti-naturalists'. Consider for example Roderick Chisholm's version of evidentialism, which is the paradigmatic anti-naturalistic position. Chisholm's view is that epistemic properties and epistemic relations are *irreducible*, meaning that they are of a kind that simply cannot be defined by a complex of psychological or familiar logical operations [Chisholm 1967]. If one looks at the dispute between methdological naturalists and Chisholmians one can see that what they have been arguing over is the place of cognitive psychology in epistemology—specifically whether a detailed causal account of human belief formation is a relevant matter to weigh in advancing a theory of justification. The point to notice is that this dispute has been conducted with a tacit agreement within the field that Chisholm was at least right about logic offering traditional epistemologists little theoretical advantage in the analysis of epistemic concepts and, more importantly, epistemic relations.

It is precisely this Chisholmian view that logic plays only a minimal analytical role in epistemology that should be abandoned. While ready-made solutions to the Gettier problem are not to be found in the journals of artificial intelligence and non-trivial conceptual and methodological issues remain in identifying and representing *relata*, we nevertheless see a role for logical AI in the very heart of traditional analytic epistemology: to analyze and model epistemic relations.

One of our interests is to see epistemologists incorporate definitions of epistemic relations, particularly inference relations, into their theories of knowledge. We think that adopting this practice would yield better theories of knowledge, which is of intrinsic interest. But adopting this practice would also be of interest to the field of knowledge representation and reasoning. For there is an emerging area of research encompassing epistemology and logical AI [Ford *et. al.* 1995; Pereira 2002], one that is created by shared interests between these two fields—shared in so far as an aim of theoretical AI is the study of the class of possible epistemic relations, the primary aim of epistemology is the specification of those properties and relations necessary to assemble a comprehensive account of knowledge, and an aim of practical AI is engineering artificial intelligence technologies that perform increasingly sophisticated inference operations on data structures.

2 Mechanizing Logic

The project of effecting logical reasoning by computers involves two fundamental notions underpinning our discussion so far, namely *defeasibility* and *constructivity.* The notion of defeasiblity, that of drawing an inference that may be undermined by additional information, figures in the move away from Cartesian foundationalism and is a notion that has been studied extensively within logical artificial intelligence: is there a principled way of making a non-monotonic inference or are all instances of non-monotonic reasoning simply too context dependent to admit a *logical* structure? Constructivity figures in logical AI because what is needed are concrete witnesses of proofs: it is not enough to know that a proposition A follows but that we have a path for the program to follow to A. But once we set about mechanizing logic by means of proof objects that behave non-monotonically, the case against thinking logical consequence is entirely distinct from inference relations begins to unravel.

Still, care must be taken in drawing the relationship between AI and epistemology. One part necessary for doing so is showing that there is a sufficient theoretical understanding of non-monotonic logics to consider applying them to the theory of knowledge. What we propose to do here is present a sketch of the correspondence results that hold between semantics for logic programs and a particular kind of logic for defeasible reasoning, called default logic. It remains to be shown that default logic is a suitable framework for representing some epistemic relation or other. We will take up this issue in another paper.

To begin, we note that a property shared by both epistemic relations and causal relations that distinguishes both from logical implication is that the former pair are uni-directional in the sense that there is no implicit contraposition. This directionality of epistemic and causal relations is an essential feature of logic programs [Colmerauer et. al. 1973; Kowalski 1974, 1979; Warren and Pereira 1977], where premises must be true in order to apply an inference rule.

An logic program P is a finite set of rules of the form,

$$C \leftarrow P_1, ..., P_n, \neg N_1, ..., \neg N_m$$

where in order to produce a result or conclusion C what is needed is a set of conditions $P_1, ..., P_n$ where each P_i is true in the program along with absence or negation of a set of negative conditions $\neg N_1, ..., \neg N_m$ where each $\neg N_i$ denotes a condition that, if satisfied, would be sufficient to prevent concluding C with this rule. As noted, the functor $\leftarrow$ does not presume explicit contraposition. Rather, we view programming clauses as expressing an inference rule, one that may be applied, procedurally from the 'bottom-up' to conclude C given all P_i's and no $\neg N_i$'s, or 'top-down' by trying to prove the body of the rule to yield C.

What we propose to do is introduce readers to enough of the semantics of logic programming to establish correspondence results between logic programs and one particular logic for defeasible reasoning, *default logic* [Reiter 1980]. The goal is to persuade readers that there now exist enough theoretical understanding of logics for defeasible inference for us to reconsider how we formulate epistemic relations.

2.1 Logic program semantics

The semantics we will present for logic programs is the extended well-founded, *WFSX*, set forth in [Alferes and Pereira 1996]. We begin by providing definitions of interpretation and model for programs extended with explicit negation.

Definition 3 (Interpretation). An *interpretation* I of a language $\mathcal{L}$ is any set $T \cup$ *not* F,[5] where T and F are disjoint subsets of objective literals over the Herbränd base, and

$$\text{if } \neg l \in T \text{ then } l \in F \text{ (Coherence Principle)}$$

where l is an objective literal. The set T contains all ground objective literals *true* in I, the set F contains all ground objective literals *false* in I. The truth value of the remaining objective literals is *undefined*.

Notice how the two types of negation become linked via the Coherence Principle: for any objective literal l, if $\neg l \in I$, then *not* $l \in I$. This definition of interpretation not only guarantees that every interpretation complies with coherence but also with noncontradiction.

Proposition 1 (Noncontradiction condition). If $I = T \cup$ *not* F is an interpretation of a program P then there is no pair of objective literals A, $\neg A$ of P such that $A \in T$ and $\neg A \in T$.

Proposition 2. Let $\mathcal{H}$ be the set of all objective literals in the language $\mathcal{L}$, $V = \{0, \frac{1}{2}, 1\}$ and $A \in \mathcal{H}$. Any interpretation $I = T \cup$ *not* F may be equivalently viewed as a function $I : \mathcal{H} \to V$, defined by:

$$I(A) = 0, \text{ if } \textit{not } A \in I;\ I(A) = 1, \text{ if } A \in I;\ I(A) = \tfrac{1}{2}, \text{ otherwise.}$$

With this function we may now define a truth valuation of formulae.

Definition 4 (Truth valuation). If I is an interpretation, the truth valuation $\hat{I} : C \to V$ where C is the set of all formulae of the language, recursively defined as follows:

- if l is an objective literal then $\hat{I} = I(l)$;
- if $s =$ *not* l is a default literal then $\hat{I} = 1 - I(l)$
- if s and r are formulae then $\hat{I}((s, r)) = min(\hat{I}(s), \hat{I}(r))$;
- if l is an objective literal and s is a formula then:

  $\hat{I}(l \leftarrow s) = 1$ if $\hat{I}(s) \leq \hat{I}(l)$ or $\hat{I}(\neg l) = 1$ and $\hat{I}(s) \neq 1$; 0 otherwise.

Definition 5 (Model). An interpretation I is called a *model of a program* P if and only if for every ground instance of a program rule $H \leftarrow B$ we have $\hat{I}(H \leftarrow B) = 1$.

Example 1. The models of the program

$$\text{P} = (\neg b;\ b \leftarrow a;\ a \leftarrow \textit{not } a,\ \textit{not } c;\ c \leftarrow \textit{not } \neg c;\ \neg c \leftarrow \textit{not } c)$$

[5]Where *not* $\{a_1, ..., a_n, ...\}$ stands for $\{\textit{not } a_1, ..., \textit{not } a_n, ...\}$.

are:

$M_1 = \{\neg b,\ not\ b\}$
$M_2 = \{\neg b,\ not\ b, c,\ not\ \neg c\}$
$M_3 = \{\neg b,\ not\ b, c,\ not\ \neg c,\ not\ a\}$
$M_4 = \{\neg b,\ not\ b,\ not\ c, \neg c\}$
$M_5 = \{\neg b,\ not\ b, \neg a,\ not\ a\}$
$M_6 = \{\neg b,\ not\ b, \neg a,\ not\ a, c,\ not\ \neg c\}$
$M_7 = \{\neg b,\ not\ b,\ not\ \neg a\}$
$M_8 = \{\neg b,\ not\ b, c,\ not\ \neg c,\ not\ \neg a\}$
$M_9 = \{\neg b,\ not\ b, c,\ not\ \neg c,\ not\ a,\ not\ \neg a\}$
$M_{10} = \{\neg b,\ not\ b,\ not\ c, \neg c,\ not\ \neg a\}$

Only M_3, M_6, and M_9 are classical 3-valued models of P, since all of the rules are true, while M_1, M_2, M_4, M_7, M_8, and M_{10} are not classical models, because in all of them the body of the rule $b \leftarrow a$ is undefined and the head is false (i.e., the truth value of the head is smaller than that of the body.). Finally, M_5 is not a classical model since in it the truth value of the head (false) of the rule $a \leftarrow\ not\ a,\ not\ c$ is smaller than that of the body (undefined).

Next we need to define stability in models, which we use to define WFSX semantics. To define the semantics, the language is expanded to include the proposition $\mathbf{u}$ such that every interpretation I satisfies $I(\mathbf{u}) = \frac{1}{2}$. In what follows a 'non-negative' program is a program whose premises are either objective literals or $\mathbf{u}$.

Definition 5 (P modulo I ($\frac{P}{I}$) transformation). Let P be an extended logic program and let I be an interpretation. P modulo I, $\frac{P}{I}$, is the program obtained from P by performing in the sequence the following four operations:

1. Remove from P all rules containing a default literal $l =\ not\ A$ such that $A \in I$;

2. Remove from P all rules containing in the body an objective literal l such that $\neg l \in I$;

3. Remove from all remaining rules of P their default literals $l =\ not\ a$ such that $not\ A \in I$.

4. Replace all the remaining default literals by proposition $\mathbf{u}$.

The resulting program is $\frac{P}{I}$ is by definition non-negative.

Definition 6 (Least operator). Let P be a non-negative program. The operator $least(P)$ is the set of literals $T \cup\ not\ F$ obtained by:

- Let P' be the non-negative program obtained by replacing in P every non-negative objective literal $\neg l$ by a new atomic symbol, '$\neg\_l$'.

- Let $T' \cup not\ F'$ be the least 3-valued model of P'.

- $T \cup not\ F$ is obtained from $T' \cup not\ F'$ by reversing the replacements above.

The least 3-valued model of a non-negative program can be defined as the least fixpoint of the following generalization of the van Emden-Kowalski least model operator Ψ for definite logic programs:

Definition 7 (Ψ^* operator). Suppose that P is a non-negative program, I is an interpretation of P and A and the A_i are all ground atoms. Then $\Psi^*(I)$ is a set of atoms defined as follows:

- $\Psi^*(I)(A) = 1$ if and only if there is a rule $A \leftarrow A_1, ..., A_n$ in P such that $I(A_i) = 1$ for all $i \leq n$.
- $\Psi^*(I)(A) = 0$ if and only if for every rule $A \leftarrow A_1, ..., A_n$ there is an $i \leq n$ such that $I(A_i) = 0$.
- $\Psi^*(I)(A) = \frac{1}{2}$, otherwise.

Theorem 1 ***(3-valued least model)*** *The 3-valued least model of a non-negative program is:*

$$\Psi^{*\ \uparrow\omega}(not\ \mathcal{H})$$

Theorem 2 *$least(P)$ uniquely exists for every non-negative program P.*

Note that $least(P)$ doesn't always satisfy the conditions of non-contradiction and coherence,

Example 2. Given the program $P = (a \leftarrow;\ \neg b \leftarrow;\ \neg a \leftarrow \neg b;\ b \leftarrow \mathbf{u})$, $least(P) = \{a, \neg a, \neg b\}$ but is not an interpretation. Both non-contradiction and coherence are violated.

Example 3. Given the program $P = (a \leftarrow \neg b;\ b \leftarrow \neg b;\ \neg a)$ and the interpretation $I = \{a, \neg a, not\ \neg b\}$, where $\frac{P}{I} = (a \leftarrow \mathbf{u}, b \leftarrow \mathbf{u}), \neg a)$. $least(\frac{P}{I}) = \{\neg a, not\ \neg b\}$, which although noncontradictory violates coherence.

To impose coherence when contradiction is not present, we define a partial operator that transforms any non-contradictory set of literals into an interpretation.

Definition 8. (The *Coh* operator). Let $QI = QT \cup not\ QF$ be a set of literals such that QT is the interpretation $T \cup\ notF$ such that

$$T = QT\ and\ F = QF \cup \{\neg l | l \in T\}.$$

The Coh is not defined for contradictory sets of literals.

The Coh operator is not a model of the program, however, since it does not take into account the consequences of applying the function. By generalizing this operation, we have the last piece necessary to define Stable Models and Well Founded Models.

Definition 9. (The Ψ operator). Let P be a logic program, I an interpretation, and $J = least(\frac{P}{I})$. If $Coh(J)$ exists, then $\Psi_P(I) = Coh(J)$. Otherwise $\Psi_P(I)$ is not defined.

Definition 10. (WFSX, PSM and WFM). An interpretation I of an extended program P is called a *Partial Stable Model* (PSM) of P if and only if $\Psi_P(I) = I$. The *F-least Partial Stable Model* is called the *Well-Founded Model* (WFM). The WFSX semantics of P is determined by the set of all PSMs of P.

2.2 Default logic semantics

Logic programming-default logic correspondence results hold for a restricted form of Reiter default theories, namely when the first-order component of default theories, the set W, contains only literals and the set of defaults, D, contains only restricted defaults, defaults of standard form, $\frac{\alpha:\beta}{\gamma}$, but where α, β and γ are literals.

It is well known that Reiter's default logic may have multiple extensions.

Example 4. Let $\Delta = \langle D, W \rangle$ where $D = \{\frac{c:\neg a}{b}, \frac{c:\neg b}{a}\}$ and $W = \{c\}$. The default theory Δ has two extensions:

$$E_1 = \{a, \neg b, c\}$$
$$E_2 = \{b, \neg a, c\}$$

Nevertheless, a skeptical consequence set may be defined for the default theory Δ as the set of literals that appear in every extension on Δ.

There are two approaches that relate logic programs with default theories, and which resolve the issue of multiple extensions. Well-founded semantics [Baral and Subrahmanian 1991] provides a semantics for default theories with multiple extensions.

Definition 11 (Well-founded semantics). Let $\Delta = \langle D, W \rangle$ be a default theory, and let E_Δ be Reiter's fixed point operator [Reiter 1980]. Since E_Δ is antitonic E_Δ^2 is monotonic, and thus has a least fixpoint (with respect to set inclusion in extensions). Then

- A formula F is *true* in a default theory Δ with respect to the well-founded semantics if and only if $F \in lfp(\mathsf{E}_\Delta^2)$;
- F is *false* in Δ w.r.t. the well-founded semantics if and only if $F \notin gfp(\mathsf{E}_\Delta^2)$;
- Otherwise F is said to be *unknown* or *undefined*.

This semantics is defined for all theories and is equivalent to the Well-Founded Model Semantics of van Gelder, Ross and Schlipf [van Gelder *et. al.* 1991] of normal logic programs.

This work has since been generalized by Przymusinska and Przumusinski by introducing the notion of stationary default extensions [Przymusinska and Przumusinski 1993.]

Definition 12 (Stationary extension). Given a default theory Δ, E is a stationary default extension if and only if:

- $E = \mathsf{E}^2_\Delta(E)$;
- $E \subseteq \mathsf{E}_\Delta(E)$.

Definition 13 (Stationary default semantics). Let E be a stationary extension of a default theory Δ such that:

- A formula L is *true* in E if and only if $L \in E$;
- A formula L is *false* in E if and only if $L \notin E$;
- Otherwise a formula L is said to be *undetermined* or *undefined*.

We note that every default theory has at least one stationary default extension. The least stationary default extension always exists, and corresponds to the well-founded semantics for default theories. Moreover, the least stationary default extension can be computed by iterating the operator E^2_Δ.

There are some properties that a default theory semantics should have. We turn to these next.

Uniqueness of minimal extensions: We say that a default theory has the *uniqueness of minimal extensions* property if when it has an extension it has a minimal one.

It is well known that Reiter's default theories do not have the uniqueness of minimal extensions property. But by obeying this property, a default semantics eases finding iterative algorithms to compute skeptical (cautious) versions of a default semantics.

Definition 14 (Union of Theories). The union of two default theories

$\Delta_1 = \langle D_1, W_1\rangle$ and $\Delta_2 = \langle D_2, W_2\rangle$ with languages $L(\Delta_1)$ and $L(\Delta_2)$ is the theory:

$\Delta = \Delta_1 \cup \Delta_2 = (D_1 \cup, D_2, W_1 \cup W_2)$ with language $L(\Delta) = L(\Delta_1 \cup L(\Delta_2)$.

Modularity. Let Δ_1 and Δ_2 be two default theories with consistent extensions such that $L(\Delta_1) \cap L(\Delta_2) = \{\}$ and let $\Delta = \Delta_1 \cup \Delta_2$, with extensions $E^i_{\Delta_1}$, $E^j_{\Delta_2}$ and E^k_Δ. A semantics for default theories is *modular* if and only if:

$\forall \mathrm{A}(\forall_i A \in E^i_{\Delta_1} \Rightarrow \forall_k A \in E^k_{\Delta_1})$

$\forall \mathrm{A}(\forall_j A \in E^j_{\Delta_2} \Rightarrow \forall_k A \in E^k_{\Delta_1})$

Informally, a default theory semantics is modular if any theory resulting from the union of two consistent theories with disjoint language contains the consequences of each of the theories alone. We remark that Reiter's default logic is modular (for a proof, see [Alferes and Pereira 1996, p. 89]).

Example 5. Consider the two default theories:

$$\Delta_1 = \left\langle \left\{ \frac{:\neg a}{\neg a}, \frac{:a}{a} \right\}, \{\} \right\rangle$$
$$\Delta_2 = \left\langle \left\{ \frac{:b}{b} \right\}, \{\} \right\rangle$$

Classical default theory, well-founded semantics, and stationary semantics all identify $\{b\}$ as the single extension of Δ_2. Since the languages of the two theories are disjoint, one would expect their union to include b in all its extensions. However, both the well-founded semantics as well as the least stationary semantics give the value undefined to b in the union theory; therefore, they are not modular. There is a conflict in the interaction among the default rules of both theories. Reiter's classical default theory is modular but returns two extensions, $\{\neg a, b\}$ and $\{a, b\}$, and thus fails to give a unique minimal extension to the union.

We say that a default rule d is *applicable* in an extension E if and only if $\alpha \subseteq E$ and $\neg\beta \cap E = \{\}$, and an applicable default is *applied* if and only if $\alpha \in E$.

Enforcedness. Given a theory Δ with extension E, a default d is *enforceable* in E if and only if $\alpha \in E$ and $\beta \subseteq E$. An extension is $enforced$ if all enforceable defaults in D are applied.

Whenever E is an extension, if a default is enforceable then it must be applied. Note that an enforceable default is always applicable. Another way of viewing enforcedness is that if the default d is an enforceable default, and E is an extension, then the default rule d must be understood as an inference rule $\alpha, \beta \rightarrow \gamma$ and so $\gamma \in E$ must hold.

Based on the notion of enforcedness, Przymusinka and Przymusinki define the notion of saturated default theories.

Definition 15 (Saturated Default Theory). A default theory $\Delta = \langle D, W \rangle$ is *saturated* if and only if for every default rule $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma} \in D$, if $\alpha \in W$ and $\beta_i \subseteq W$, for $1 \leq i \leq n$, then $\gamma \in W$.

For this class of default theories Przymusinka and Przymusinki prove that both stationary and well founded default semantics comply with enforcedness. However, considering only saturated default theories is a significant restriction: all conclusions of the defaults are already in the W component of the theory.

We are now close to presenting the correspondence theorem between logic programs and default theories. In order to relate default theories to extended logic programs, however, we must provide a modular semantics for default theories. Therefore, we now present a modular and enforced semantics for a class of default theories called Ω-default theories.

2.3 Ω-default theory

In this section we present a default theory semantics that is modular and enforced for every restricted default theory. Moreover, when it is defined it has a unique minimal extension.

To link default theories to extended logic programs, we must provide a modular semantics in the case of contradictory default theories.

Example 6. In the default theory:

$$\left\langle \left\{ \frac{:}{\neg a} \right\}, \left\{ \frac{:}{a} \right\}, \{\} \right\rangle$$

its two default rules with empty prerequisites and justifications should always be applied, which clearly enforces a contradiction. Note that this would also be the case in the default theory $\langle \{\}, \{a, \neg a\} \rangle$.

Reconsider now Example 5, which demonstrates that stationary default semantics are non-modular, where $D = \{\frac{:\neg a}{\neg a}, \frac{:a}{a}, \frac{:b}{b}\}$ and $\{\}$ is the least stationary extension.

This result is obtained because $\mathsf{E}_\Delta(\{\})$, by having $\neg a$ and a, forces, via the deductive closure, $\neg b$ (and all the other literals) to belong to it. This implies the non-applicability of the third default, $\frac{:b}{b}$, in the second iteration. For that not to happen one should inhibit $\neg b$ from belonging to $\mathsf{E}_\Delta(\{\})$, which can be done by preventing the trivialization by inconsistency generated by the deductive closure condition of the operator E. We avoid this problem in a logic programming context, since formulae of logic programs are just literals. We may simply rename negative literals. We now incorporate this idea into the definition of the fixed-point operator E'_Δ.

Definition 16 ($\mathsf{E}'_\Delta(E)$). Let $\Delta = \langle D, W \rangle$ and E be an extension. Let E' be the smallest set of atoms which:

1. contains W';
2. is closed under all derivation rules of the form $\frac{\alpha:\beta}{\gamma}$, such that

   $\frac{\alpha:\beta}{\gamma} \in D$, and $\neg f \notin E$, for every $\ulcorner \neg f \in \beta' \urcorner$, and $f \notin E$ for every $\ulcorner not\_f \in \beta' \urcorner$.

   where the new W', α', β', and γ' are obtained from the original W, α, β, and γ by replacing every negative literal $\ulcorner \neg\varphi \urcorner$ in the originals by a new atom $\ulcorner not\_\varphi \urcorner$.

   $\mathsf{E}'_\Delta(E)$ is obtained from E' by replacing every atom of the form $\ulcorner not\_\varphi \urcorner$ by $\ulcorner \neg\varphi \urcorner$.

Definition 17 (Semi-normal default theories). Given a default theory Δ, its semi-normal version Δ^{sem} is obtained by replacing each default rule $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$ in Δ by the default rule:

$$s^{sem} = \frac{\alpha : \beta_1, \ldots, \beta_n, \gamma}{\gamma}.$$

We now turn to defining the Ω_Δ fixed-point operator, Ω-extensions, and the Ω-default semantics.

Definition 18 (Ω_Δ operator). For a theory Δ we define:

$$\Omega_\Delta(E) = \mathsf{E}'_\Delta(\mathsf{E}'_{\Delta^{sem}}(E)). \ \Box$$

Definition 19 (Ω-extension). Let Δ be a default theory. E is an extension if and only if

- $E = \Omega_\Delta(E)$
- $E \subseteq \mathsf{E}'_{\Delta^{sem}}(E)$.

Given the notion of Ω-extensions, we may now define the semantics for a default theory.

Definition 20 (Ω-default semantics). Let Δ be a default theory. E is an extension on Δ, and l a literal.

- l is *true* w.r.t. extension E if and only if $l \in E$;
- l is *false* w.r.t. extension E if and only if $l \notin \mathsf{E}'_{\Delta^{sem}}(E)$;
- Otherwise l is *undefined*.

The Ω-default semantics of Δ is determined by the set of all Ω-extensions of Δ. The *skeptical* (or *cautious*) semantics of Δ is determined by the least Ω-extensions of Δ, whose existence are guaranteed by the uniqueness of minimal extensions theorem below.

But noting that a default theory Δ is contradictory if and only if it has no Ω-extension, we may prove that the Ω-default semantics has the three properties mentioned above—*uniqueness of minimal extensions, modularity,* and *enforcedness*—as necessary to establishing correspondence between logic programs and default logic. All three theorems and their proofs appear in [Alferes and Pereira 1996].

Theorem 3 *(Uniqueness of minimal extensions) If Δ has an extension then there is one least extension E.*

Theorem 4 *(Enforcedness) If E is an Ω-extension then E is enforced.*

Theorem 5 *(Modularity) Let L_{Δ_1} and L_{Δ_2} be the languages of two default theories. If $L_{\Delta_1} \cap L_{\Delta_2} = \{\}$ then, for any corresponding extensions E_1 and E_2, there always exists an extension E of $\Delta = \Delta_1 \cup \Delta_2$ such that $E = E_1 \cup E_2$.*

2.4 Correspondence between logic programs and default theories

We may now state the equivalence of Ω-extensions and partial stable models of extended logic programs as defined above. For proofs, the reader is again referred to [Alferes and Pereira 1996].

Definition 21 (Program correspondence to a default theory). Let $\Delta = \langle D, \{\} \rangle$ be a default theory. We say than an extended logic program P corresponds to Δ if and only if:

- For every default of the form $\frac{\alpha_1,...,\alpha_n:\beta_1,...,\beta_m}{\gamma} \in \Delta$ there exists a rule $\ulcorner \gamma \leftarrow \alpha_1, ..., \alpha_n, not\neg\beta_1, ..., not\neg\beta_m \urcorner \in P$, where $\neg b_j$ denotes the $\neg$-complement of b_j.
- no rules other than these belong to P.

Definition 22 (Interpretation corresponding to a context). An interpretation I of a program P corresponds to a default context E of the corresponding default theory T if and only if for every objective literal l of P (and literal l of T):

- $I(l) = 1$ if and only if $l \in E$ and $l \in \mathsf{E}'_{\Delta^{sem}}(E)$
- $I(l) = \frac{1}{2}$ if and only if $l \notin E$ and $l \in \mathsf{E}'_{\Delta^{sem}}(E)$
- $I(l) = 0$ if and only if $l \notin E$ and $l \notin \mathsf{E}'_{\Delta^{sem}}(E)$.

We note that Reiter default theories are a generalization of restricted default theories in the sense that whenever Reiter semantics (E-extension) assigns a meaning to a theory (i.e., the theory has at least one E-extension), Ω-semantics assigns one also.

Theorem 6 *(Correspondence) Let $\Delta = \langle D, \{\} \rangle$ be a default theory corresponding to program P. E is an Ω-extension of Δ if and only if the interpretation I corresponding to E is a partial stable model of P.*

So, according to this theorem we can say that explicit negation is nothing but classical boolean negation in (restricted) default theories, and *vice-versa*. What this theorem allows us to do is to rely on the top-down procedures of logic programming to compute default extensions—that is, this provides us with a sound procedure for Reiter's default logic.

3 Conclusion

To recap, what we've done is to discuss the state of contemporary epistemology and argue that more attention should be paid to formulating epistemic relations within theories of knowledge. We suggested that the logical AI literature offers a theoretical underpinning for such an effort, primarily because the theoretical constraints imposed by mechanizing logic address key obstacles faced when considering the project of modeling defeasible inference relations. We then illustrated how this is so by briefly discussing a semantics for default theories and logic programs, providing enough of a description to sketch how the correspondence results are obtained.

Acknowledgement 7 *This research was supported, in part, by grant SES 990-6128 from the National Science Foundation, by a postdoctoral scholarship from CENTRIA, and by POCTI project 40858 "FLUX - FleXible Logical Updates".*

References

[1] Alferes, J. A., C. Damásio and L. M. Pereira. 1995. "A logic programming system for non-monotonic reasoning," *Journal of Automated Reasoning* 14: 93-147.

[2] Alferes, J. A., L. M. Pereira and T. Przymusinska. 1998. "'Classical' negation in non-monotonic reasoning and logic programming," *Journal of Automated Reasoning* 20: 107-142.

[3] Alferes, J. A. and L. M. Pereira. 1996. *Reasoning with Logic Programming*, Berlin: Springer-Verlag.

[4] Baral, C. and V. S. Subrahmanian. 1991. "Dualities between alternative semantics for logic programming and non-monotonic reasoning," in A. Nerode, W. Marek and C. S. Subrahmanian (eds.), *LP & NMR*, Cambridge: MIT press, pp. 69-86.

[5] Bidoit, N. and C. Froidevaux. 1988. "General logic databases and programs: default logic semantics and stratification," *Journal of Information and Computation.*

[6] Chisholm, R. 1966. *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall.

[7] Clark, K. 1978. "Negation as Failure," in Gallaire, H. and J. Minker, [eds.] *Logic and Data Bases*, 293-322.

[8] Colmerauer, A., et. al. 1973. "Un Système de Communication Homme-Machine en Français," Research Report. France: Université Aix-Marseille II, Groupe d'Intelligence Artificielle.

[9] Conee, E. 1987. "Evident, but Rationally Unacceptable," *Australasian Journal of Philosophy*, 65: 316-26.

[10] Cross, C. 2000. "The Paradox of the Knower without Epistemic Closure," *Mind* 110: 329-332.

[11] Damásio, C. and L. M. Pereira. 2001. "Antitonic logic programs," in *Procs. 6th Int. Conference on Logic Programming and Non-monotonic Reasoning* (LPNMR '01), T. Eiter and M. Truszczynski (eds.), Springer LNAI 2001.

[12] De Rose, K. and T. Warfield. 1999. *Skepticism: A Contemporary Reader*, Oxford: Oxford University Press.

[13] Feigenbaum, E. and J. Feldman [eds.]. 1963. *Computers and Thought.* New York: McGrall Hill.

[14] Feldman, R. 1974. "An Alledged Defect in Gettier Counterexamples," *Australasian Journal of Philosophy* (52): 68-69.

[15] Ford, K., C. Glymour and P. Hayes, [eds.] 1995. *Android Epistemology*, Cambridge: MIT Press.

[16] Gettier, E. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121-123.

[17] Green 1969a. "Application of Theorem Proving to Problem Solving," *Proceedings of the First International Joint Conference on Artificial Intelligence*, Washington, D.C. Los Altos, CA: Morgan Kaufmann, 219-239.

[18] Green 1969b. "Theorem-Proving by Resolution as a Basis for Question-Answering Systems," appearing in Meltzer, B. and D. Michie [eds.] *Machine Intelligence* 4. Edinburgh: Edinburgh University press. 183-205.

[19] Goldman, A. 1967. "A Causal Theory of Knowing," *Journal of Philosophy* 64: 357-372.

[20] Goldman, A. 1986. *Epistemology and Cognition*, Cambridge: Harvard University Press.

[21] Harman, G. 2001. "Internal Critique: A Logic is not a Theory of Reasoning and a Theory of Reasoning is not a Logic," appearing in *Studies in Logic and Practical Reasoning, Vol. 1.* Gabbay, D. *et. al.* [ed.]. London: Elsevier Science.

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

[23] Hoffman, R. 1991. "Human factors psychology in the support of forecasting: The design of advanced meteorological workstations," *Weather and Forecasting* (6): 98-110.

[24] Kaplan, D. and R. Montague. 1960. "A Paradox Regained," *Notre Dame Journal of Formal Logic* 1: 79-90.

[25] Kim, J. 1988. "What is 'Naturalized Epistemology'?" *Philosophical Perspectives 2*, J. Tomberlin [ed.]. Atascadero, CA: Ridgeview Publishing, 381-406.

[26] Kowalski, R. 1974. "Predicate logic as a programming language," *Proceedings IFIP'74*, Amsterdam: North Holland Publishing, 569-574.

[27] Kowalski, R. 1979. "Algorithm = logic + control," *Communications of the ACM*, 22: 424-436.

[28] Kyburg, H. E., Jr. 1961. *Probability and the Logic of Rational Belief.* Middletown, CT: Wesleyan.

[29] Kyburg, H. E., Jr. 1997. "The Rule of Adjunction and Rational Inference," *Journal of Philosophy* 94:109-25.

1999. "Statistical Inference as Default Logic," *International Journal of Pattern Recognition and Artificial Intelligence* 13(2) : 267-283.

[30] Makinson, D. 1965. "The paradox of the preface," *Analysis* 25: 205-7.

[31] Meyer, J. J. and W. van der Hoek. 1995. *Epistemic Logic for AI and Computer Science*, Cambridge: Cambridge University Press.

[32] Pereira, L. M. 2002. "Philosophical Incidence of Logic Programming", in D. Gabbay *et. al.* (eds), *Studies in Logic and Practical Reasoning*, Vol. 1. Elsevier Science, pp. 425-448.

[33] Prior, J. 2001. "Highlights of Recent Epistemology" *The British Journal for the Philosophy of Science* 52: 1-30.

[34] Przymusinka, H. and T. Przymusinki 1993. "Stationary default extensions," Technical Report, Department of Computer Science, California State Polytechnic and Department of Computer Science, University of California at Riverside.

[35] Quine, W. V. 1969. *Ontological Relativity and Other Essays*, New York: Columbia University Press.

[36] Reiter, R. 1980. "A Logic for Default Reasoning," *Artificial Intelligence*, 13:81-132.

[37] Sellen, A. and R. Harper. 2001. *The Myth of the Paperless Office*, Cambridge: MIT Press.

[38] Uzquiano, G. Forthcoming. "The Paradox of the Knower without Epistemic Closure?".

[39] van Emden, M. and R. Kowalski. "The semantics of predicate logic as a programming language," *Journal of ACM* 4(23): 733-742.

[40] van Gelder, A. and K. A. Ross and J. S. Schlipf. 1991. "The well-founded semantics for general logic programs," *Journal of ACM*, 38(3): 620-650.

[41] Warren, D. H. D. and L. M. Pereira and F.C.N. Pereira. 1977. "PROLOG—The Language and Its Implementation Compared with LISP," *Proceedings of the Symposium on Artificial Intelligence and Programming Languages, SIGPLAN Notices*, 12(8) and *SIGART Newsletter* (64): 109-115.

[42] Wheeler, G. R. 2002. "Kinds of Inconsistency", appearing in Carnelli, et, al. [ed]., *Paraconsistency*, New York: Marcel Dekker.

[43] Wheeler, G. R. "Statistical Default Logic," forthcoming.