Epistemology and 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

May 17, 2004

Abstract

In this essay we advance the view that analytical epistemology and artificial intelligence are complementary disciplines. Both fields study epistemic relations, but whereas artificial intelligence approaches this subject from the perspective of understanding formal and computational properties of frameworks purporting to model some epistemic relation or other, traditional epistemology approaches the subject from the perspective of understanding the properties of epistemic relations in terms of their conceptual properties. We argue that these two practices should not be conducted in isolation. We illustrate this point by discussing how to represent a class of inference forms found in standard inferential statistics. This class of inference forms is interesting because its members share two properties that are common to epistemic relations, namely *defeasibility* and *paraconsistency*. Our modeling of standard inferential statistical arguments exploits results from both logical artificial intelligence and analytical epistemology. We remark how our approach to this modeling problem may be generalized to an interdisciplinary approach to the study of epistemic relations.

[Keywords: statistical default logic; non-monotonic reasoning; epistemic closure; logic programming; uncertainty; knowledge representation.]

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. 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 2004] and the paradoxes of rational acceptance, namely the lottery [Kyburg 1961, 1997] and the preface [Makison 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. Get-

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

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

tier'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 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

[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 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].

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 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 a 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.[5] 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

[5]Although our discussion so far has been in terms of traditional epistemology, we should stress that our view that logical AI is relevant to the study of epistemic relations only depends upon a theory discussing some epistemic property, like justification, and that its means of propagation behave sufficiently different than closure under logical consequence. Thus our thesis is quite independent of various approaches within traditional epistemology. Our thesis is also compatible with the recent proposal [Williamson 2000] to reverse the direction of explanation within epistemology, by denying that notions like 'belief' and 'justification' are conceptually more basic than knowledge and instead treat the concept of knowledge itself as basic and thus a necessary constituent of an analysis of justification.

et. al. 1995; Ello 2002, 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.

To motivate philosophers to take a closer look at the tools available in logical AI and also to motivate computer scientists to take a closer look at what problems contemporary epistemologists are working on, we will discuss in this essay how to represent a class of inference forms found in standard inferential statistics. The reason that we will concentrate on this modeling problem is that it features two key properties that are important to understand when modeling epistemic relations, *defeasibility* and *paraconsistency*. Standard statistical inference has proved stubbornly resistant to logical analysis, much like defeasible closure conditions in epistemology. Also, constructing an inference relation with this property forces us to think more carefully about consistency and coherence conditions. Once we have this representation scheme we will then discuss how to test the behavior of these relations within logic programming and then discuss the prospect of transforming mechanized epistemic relations into epistemic tools. What follows then may be thought of as an exercise in the study of epistemic relations.[6]

2 Defeasibility and Non-monotonic Inference

The notion of defeasibility figures in the discussion of epistemic relations and also in logical artificial intelligence. An inference to a proposition A is defeasible if additional information added to the premises undermines that inference. The limiting case is learning that A is in fact false, which would signal that something is amiss with making an inference to A based upon this set of premises. More interesting cases of defeasibility arise when an inference to a claim is undermined by additional, non-contradictory information. For instance, the mean height of a sample of high school students drawn at random is typically reasonable grounds for concluding that the mean height of the sample is a close estimate of the mean height of the school's student body. Standard statistical inferences such as this one are defeasible precisely because we may learn new, non-contradictory information that undermines the support for thinking that the sample is representative. For instance, if we were to learn that the random sample drawn is in fact composed exclusively of the members of the varsity basketball team, then we would no longer consider the sample as likely being a close estimate of the mean height of the student body.

[6]It is important to note that one may accept our call to formally define epistemic relations without accepting either a computational view of mind or the view that logic provides "the rules of thought." Such descriptive views about mind and rationality are independent of what we advance in this essay.

One way to represent this type of defeasibility is in terms of a logic that has a genuinely non-monotonic consequence relation—one whereby premises may increase in number while the number of conclusions may decrease.[7] What we propose to do in this section is to present the structure of this statistical inference in terms of a particular kind of non-monotonic inference rule, called a *statistical default*, which is used within *statistical default logic* [Wheeler 2004] to represent a variety of argument forms common in standard inferential statistics.

Let us return to the high school example. Statistical default logic suggests both an analysis of the logical structure of individual statistical inferences—such as that involved in estimating a high school class's mean height—and also provides a scheme for representing arguments composed of a sequence of statistical and deductive inference steps. Statistical default logic is a variation of Ray Reiter's *default logic* [Reiter 1980]. Default logic is a non-monotonic logic formed by augmenting first-order classical logic with non-monotonic inference rules, called *defaults*, that appear in the object language. Let α, γ and β_i's be wffs in the first-order language. Defaults are inference rules of the form

$$\frac{\alpha : \beta_1, ..., \beta_n}{\gamma}, \tag{1}$$

interpreted roughly to mean that given α and the absence of any negated β_i's, conclude γ by default. The β_i's in (1) correspond to conditions the *absence* of which, when α holds, allows γ to be inferred. The non-monotonic behavior of defaults rests in the possibility that one of the default justifications that permits the rule to be applied may be triggered by new information, thus blocking the applicability of that rule.[8]

It turns out there is a structural similarity between the workings of default rules and a class of standard statistical inference forms, of which estimating the mean height of a student body is an instance. In making a statistical inference the aim is to select a sample that represents the population with respect to some specified parameter. Often this is achieved by a series of tests designed to detect bias in the sample. It was first noticed in [Kyburg and Teng 1999] that in making a statistical inference, some conditions are satisfied explicitly, like premises, while others behave like default justifications. Typically a sample is regarded representative of a population when a few explicit conditions hold (like that the sample be drawn from the target population and the distribution of error is normal) and when it is not known that the sample is biased, which

[7] Contrast this with probability functions, which *qua* mathematical functions are monotonic: $A \subseteq B \Rightarrow Pr(A) \leq Pr(B)$. If A is a smaller part of the sample space than B, then A must be less than or equal to the probability of B. Note that the same holds for conditional probability functions in their first position, but of course not in their second: the probability of A may remain constant, increase or decrease when conditioned on a smaller part of the sample space. Nevertheless, probability is inferentially monotonic: probability premises yield probability conclusions and a superset of those premises yields the same conclusions or a superset of them [Kyburg 2001].

[8] Reiter's original paper [Reiter 1980] offers a comprehensive introduction to default logic. A good textbook treatment of default logic is [Marek and Truszczyński 1991]. We will set forth a semantics for default logic in section 4.

translates to the absence of information that would suggest a biased sample. These latter conditions express weaker assumptions than explicitly holding that the sample is representative, for if we could help ourselves to making that explicit assumption we wouldn't need inferential statistics. The underlying point is that we don't need to accept that the sample is unbiased but rather have no reason, given what we already accept, to infer that the sample is biased.

But defaults provide only half of the structure of a statistical inference since there isn't a means within the logic to distinguish between rules that rigorously probe for error and rules that let nearly any sample skate by. This is just to say that another important feature of standard statistical inference is its emphasis on the control of error. In making statistical inferences one accepts a conclusion along with a warning that there is a small, preassigned chance that the conclusion is false. A statistical inference controls error to the extent that its advertised frequency of error corresponds *in fact* to the chance one faces in making that inference and its conclusion being false. What is problematic about representing inferential statistical forms in terms of defaults is that there is no means to represent the error-probabilities of each statistical inference.

S-defaults differ from defaults by explicitly representing the *upper limit* of the s-default's probability of error.[9] Call a default in the form of

$$\frac{\alpha : \beta_1, ..., \beta_n}{\gamma}\epsilon, \tag{2}$$

an ϵ-*bounded statistical default* and the upper limit on the probability of error-parameter ϵ an ϵ-*bound* for short, where $\frac{\alpha:\beta_1,...,\beta_n}{\gamma}$ is a Reiter default and $0 \leq \epsilon \leq 1$. The schema (2) is interpreted to say that provided α and no negated β_i's, the probability that γ is false is no more than ϵ. (A Reiter default is a limiting case of a statistical default, namely when $\epsilon = 0$). A statistical default is sound just when the upper limit of the probability of error is *in fact* ϵ. An s-default is a good inference rule if it is sound and ϵ is relatively small, typically less than 0.05.

A statistical default theory is analogous to a default theory[10], except that the pair consists of a set of bounded sentences, rather than a set of closed first-order formulae, and a countable set of s-defaults, rather than a countable set of defaults.

Definition 1. A *statistical default theory* Δ_s is an ordered pair $\langle W, S\rangle$, where W is a set of bounded sentences, and S a set of statistical defaults.

A bounded sentence is a sentence-real number pair, $\langle \phi, \epsilon\rangle$ or $(\phi)_\epsilon$ for short, where ϕ is a *wff* from a first-order language $\mathcal{L}$ and $\epsilon \geq 0$. We stipulate that $(\phi)_\epsilon \equiv \phi$ when $\epsilon = 0$ and will make use of a function, $Crop(X)$, that takes

[9] A trivial corollary of the probability of error $\hat{\alpha}$ for a statistical inference is the upper limit of the probability of error, denoted by ϵ. So, if $\hat{\alpha} = 0.05$ is understood to mean that the probability of committing a Type I error is 0.05, then $\epsilon = 0.05$ is understood to mean that the probability of committing a Type I error is no more than 0.05.

[10] A default theory is a pair $\langle D, W\rangle$ where D is a (countable) set of defaults and W is a set of closed first-order formulae.

as arguments a set X of bounded sentences and returns the set of first-order formulae that appears in the first position of every pair in X.

The error-bound parameter ϵ is a guarantee that the frequency of error does not exceed ϵ. This condition complicates the two kinds of closure operations that appear in statistical default logic, since and inference to ϕ that is bounded by ϵ and another inference to ψ that is bounded by ϵ is no guarantee that $\phi \wedge \psi$ is bounded by ϵ. The next three theorems show that a conclusion appears as a statistical conclusion only if it is the result of a chain of statistical default inferences that is within the designated error bound (theorem 1), the result of a chain of deductive inference steps that is within the designated error bound (theorem 2) or if it appears within a statistical extension (definition 2), which is constructed only from inference chains of mixed type whose results are bounded by the designated error bound (theorem 3). For details and proofs, the reader is referred to [Wheeler, forthcoming].

Theorem 1 (Wheeler 2004) Let S be a set of statistical defaults, Π a set of bounded sentences, $(\gamma)_{\epsilon_\gamma}$ a bounded sentence and $Sn_\epsilon(\Pi)$ be the s-default closure of Π under S within ϵ. Define a statistical default inference chain on Π within ϵ as a sequence of bounded sentences, $\langle (\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}} \rangle$, such that $(\phi_i)_{\epsilon_{\phi_i}}$ is an ϵ-bounded conclusion from $\Pi \cup \{(\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_{i-1})_{\epsilon_{\phi_{i-1}}}\}$, where $1 \leq i \leq n$. If $(\gamma)_{\epsilon_\gamma} \in Sn_\epsilon(\Pi)$, then there is an s-default inference chain $\langle (\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}, (\gamma)_{\epsilon_\gamma} \rangle$ on Π that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion.

Theorem 2 (Wheeler 2004) Let Π be a set of bounded sentences, $(\gamma)_{\epsilon_\gamma}$ a bounded sentence and $Cn_\epsilon(\Pi)$ be the ϵ-bound closure of Π. Define a deductive inference chain as a sequence of ϵ-bounded sentences, $\langle (\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_n)_{\epsilon_{\psi_n}} \rangle$ such that $(\psi_i)_{\epsilon_{\psi_i}}$ is an ϵ-bounded consequence of $\Pi \cup \{(\psi_1)_{\epsilon_{\psi_1}}, ..., (\psi_{i-1})_{\epsilon_{\psi_{i-1}}}\}$, where $1 \leq i \leq n$. If $(\gamma)_{\epsilon_\gamma} \in Cn_\epsilon(\Pi)$, then there is a deductive inference chain $\langle (\phi_1)_{\epsilon_{\phi_1}}, ..., (\phi_n)_{\epsilon_{\phi_n}}, (\gamma)_{\epsilon_\gamma} \rangle$ of deductions on Π that yields $(\gamma)_{\epsilon_\gamma}$ as an ϵ-bounded conclusion.

Definition 2. Where $\Delta_s = \langle W, S \rangle$ at ϵ is a statistical default theory and Π is some set of bounded sentences, let $\mathsf{E}_{\Delta_S}(\Pi)$ be a minimal set satisfying three conditions:

[SD1.] $W \subseteq \mathsf{E}_{\Delta_S}(\Pi)$.
[SD2.] $Cn_\epsilon(\mathsf{E}_{\Delta_S}(\Pi)) = \mathsf{E}_{\Delta_S}(\Pi)$.
[SD3.] $\mathsf{E}_{\Delta_S}(\Pi)$ is closed under S within ϵ, i.e. for all $\frac{(\alpha)_{\epsilon_\alpha} : (\beta_1)_{\epsilon_1}, ..., (\beta_n)_{\epsilon_n}}{\gamma} \epsilon_s \in S$, $(\alpha)_{\epsilon_\alpha} \in \mathsf{E}_{\Delta_S}(\Pi)$, $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$, $\epsilon_\alpha + \epsilon_s = \epsilon_\gamma$ and $\epsilon_\gamma \leq \epsilon$.

A set of bounded sentences Π is a *statistical extension* for Δ_s at ϵ iff $\mathsf{E}_{\Delta_S}(\Pi) = \Pi$.

Theorem 3 (Wheeler 2004) Let Π be a set of bounded sentences, let $(\alpha)_{\epsilon_1}, (\beta)_{\epsilon_2}, (\gamma)_{\epsilon_3}, (\phi)_{\epsilon_4}$, and $(\psi)_{\epsilon_5}$ be ϵ_i-bounded counterparts to sentences $\alpha, \beta, \gamma, \phi$,and ψ in $\mathcal{L}$, and let $\Delta_S = \langle W, S\rangle$ at ϵ be a closed statistical default theory. Define

- *For all* $(\phi_i)_{\epsilon_{\phi_i}} \in W, \epsilon_{\phi_i} = 0$.
- $\Pi_0 = W$, and for $i \geq 0$,
- $\Pi_{i+1} = Cn_\epsilon(\Pi_i) \cup \{\gamma | \frac{(\alpha)_{\epsilon_\alpha} : (\beta_1)_{\epsilon_1}, ..., (\beta_n)_{\epsilon_n}}{\gamma} \epsilon_s) \in S$, where $(\alpha)_{\epsilon_\alpha} \in \Pi_i$ and $\neg\beta_1, ..., \neg\beta_n \notin Crop(\Pi)$ and $\epsilon_\alpha + \epsilon_s \leq \epsilon\}$.

Then Π is a statistical extension for Δ_S at ϵ iff $\Pi = \bigcup_{0 \leq i \leq \infty} \Pi_i$

Turning our attention to the mean height of the high school class, let 'h' denote a sample of high school students drawn from 'H', the entire high school class. We may think of this inference in terms of a s-default by making the following substitutions:

> α : The measured mean height of h is 195cm, written $m(\overline{h}) =$ 195cm, and measurement errors are distributed normally with mean zero and variance σ^2, written $((\mu - X)$ is $N(0, \sigma^2))$.
>
> γ : The mean height of H in cm is within two standard deviations of 195, written $(m(\overline{h}) = 195 - 1.96\sigma \leq \mu \leq m(\overline{h}) = 195 + 1.96\sigma)$, where '$\mu$' is the mean height of H and '1.96σ' replaces 'X'.
>
> β_1 : $m(\overline{h})$ is the only measured value we have for estimating the value of μ.
>
> β_2 : There is no prior statistical knowledge of the distribution of height in a class of cases that H belongs to that would lead to a conflicting inference.
>
> β_3 : The tape measure is calibrated.
>
> β_4 : The tape measure is applied correctly to the sample.
>
> β_5 : There is no information concerning the condition of the sample that preëmpts the information provided by the measurement.
>
> $\epsilon = 0.05$: The probability of error of this inference form does not exceed 5%.

Notice that we could collect additional measurements of the height of the students from the class, thus triggering justification β_1 and undermining the conclusion drawn from *this* rule. Surely if we have two measurements, we should use a distribution for the average of the two values (in most cases) and that uses a smaller variance. Notice that whether this, or one of the other justifications $\beta_1, ..., \beta_5$ is triggered does not undermine the prerequisite. It remains the case that the measured length of h is 195cm and that the distribution of errors for that tape measure is normal, with a mean of zero and its characteristic variance. It is the consequent, the conclusion that claims that the mean height of

the student body is 195cm $\pm 2\sigma$, that is blocked. Notice, too, that it is blocked when we have additional information about the sample.

Justification β_2 says that if there is prior statistical information regarding the mean height of the entire class, then that information should take precedence over any conclusion drawn from the sample mean. As a special case we may have the height of all students available, in which case the mean of those values should take precedence over any conclusion drawn from the mean of the sample. This point may be clearer if we reflect on what we would do if we measured a known height (e.g., a standard): we would not infer that the measurement recorded by the tape measure of the standard supersedes the value given by the standard length. When these two values disagree, our interest is to calibrate the tape measure, not infer the 'true' height of the standard. Likewise, if we already have knowledge of the height of the sample being measured, this knowledge should block the application of this particular s-default rule.

Justification β_3 could be considered as positive knowledge, rather than a default justification. But, in practice, we often don't know that our instruments are calibrated. Either we use them straight out of the box, taking the manufacturer's word for its variance properties, or schedule the equipment for periodic calibration. It is perhaps more faithful to actual practice to consider instruments calibrated in these circumstances until evidence suggests otherwise; if we're collecting strange data in the laboratory that is skewed, a valid metrology sticker isn't sufficient grounds to question the calibration of the instrument. In fact, the instrument's calibration is one of the first things called into question given unusual results. It is important once again to notice that such happenings do not affect the theory of error that appears in the prerequisite, nor do they mitigate the corrected measurement appearing in the prerequisite..

Justification β_4 concerns the relationship between the measurement report of the sample's mean and the goal of measurement, the mean height of H. We could cite a list of things that make up 'appropriate application': making sure that the end of the tape is flush against the floor; assuring that the tape is straight and taunt; and reading the take straight on rather than at an angle. Again, these conditions need not be exhaustive nor do we need to know that in fact all of the conditions were satisfied—that is, we do not need to know that in the tape was applied correctly in every measurement, and so on. Rather, we look for reasons to think these conditions false. Note also that we could represent these conditions as $\beta_4, \beta_{4'}$, and $\beta_{4''}$; but that is a bookkeeping issue. The result is the same: the distribution of error that is appropriate for making the default inference go through is not at issue when we misuse the instrument.

The last default, β_5, concerns general conditions that should be in place to get good measurement readings of height. When measured, students should stand flat-footed, not slouch, remove shoes and the like.

The error bound parameter ϵ says of this inference form that when applied it exposes you to no more than a 5% chance of the conclusion γ being false.

Making the appropriate substitutions for the terms in $(\alpha : \beta_1, ..., \beta_5/_\epsilon \gamma)$, the 0.95 confidence level non-monotonic inference rule may be expressed as:

$$\frac{m(\overline{h}) = 195cm \wedge ((\mu - X)\ is\ N(0, \sigma^2)) : \beta_1, ..., \beta_5}{m(\overline{h}) - 1.96\sigma \leq \mu \leq m(\overline{h}) - 1.96\sigma} 0.05$$

where $m(\overline{h}) = 195$ is the mean of the measurement height of the sample, in centimeters; $((\mu - X)$ is $N(0, \sigma^2))$ is that the distribution of errors of the tape measure is normal, with a mean of 0 for corrected readings and a variance of σ^2; $\beta_1, ..., \beta_5$ are the list of justifications that allow the rule to be applied with a probability of error not to exceed 5% so long as no negated β_i's may be inferred by derivation or applicable s-default relative to an s-default theory (and within a pre-assigned error-bound); and $m(\overline{h}) - 1.96\sigma \leq \mu \leq m(\overline{h}) + 1.96\sigma$ is the claim that the true mean height of H, μ, lies in the interval drawn around $m(\overline{h})$.

3 Tolerating Inconsistency

In the last section we considered an estimation example and proposed representing this inference within statistical default logic. We noted that each statistical inference form is bounded in error by ϵ. So we might ask what happens when we apply a sound s-default rule but are unlucky and commit an error?

Committing an error simply amounts to accepting a false statement. A false s-default consequent is just an accepted statement that proposes an interval (on the basis of a collection of evidence reports generated by a reliable measurement procedure) that in fact fails to contain the true value of μ.[11] It is important to notice that while erroneous, an accepted but false statement is *warranted*. The statement is introduced by the correct application of an statistical default rule: no such statements are introduced to the theory by mistake.

A reader troubled by the uncertainty introduced by s-defaults might recommend that to decrease the chances of falling into error, simply increase the distance in the interval around μ. Increasing the interval around μ increases the margin of error and thus reduces the frequency of accepted false evidence statements over the long run. Although mathematically sound, this would be bad advice to follow. Remember that the final goal isn't error elimination but finding the mean height of the class. If our interest is to hit bulls-eye, we are hardly helped by increasing our target to the barn its pinned on. Our interest in the true mean value and access only to finite trials presses us to accept a minimal interval around μ whose associated probability of error is known and small enough to ensure confidence, not certainty. Error is but one parameter in this optimization problem and is eliminated completely only at the price of triviality.[12]

[11] Another kind of error arises by *failing to accept* that the value of μ is within the acceptance interval when the true value is in fact within that interval.

[12] At extremes one could say that the height of μ is a real number. But that tells you only that μ is a magnitude.

What is interesting about s-default consequents is that a collection of them can be inconsistent [Wheeler 2002]. Such measurement statements are approximations at best, which is the reason they are interval-valued. However, these interval-valued statements are also fallible. S-defaults tell us that, over the long run, a small ϵ proportion of consequents will be accepted yet false. It is important to notice that accepting a false evidence statement in this manner is an error and not necessarily an inconsistency. But an interesting case is when there are sufficiently many evidence statements, one million for example, that are generated by some s-default where each statement is accepted at a 99% confidence level. A 0.99 confidence level s-default entails that after applying that s-default one million times we can be 99% sure that at least 500 of the accepted statements are false. So, even though each of the 1 million statements are accepted as true with 0.99 confidence we also accept with no less confidence that at least 500 are false.

The idea of an inference procedure that builds in inconsistency is likely to meet strong resistance. It might be thought, for instance, that the advice to measure twice and record once holds at least the promise of eliminating error, thereby avoiding the problem of accumulating inconsistent statements. Notice that what this suggestion to measure twice–write once amounts to is simply to run an experiment. With multiple measurement 'trials' we can expect to catch the very kind of errors under discussion. Unless one is a systematic incompetent, one could discover false reports by repeating the measurement procedure and tossing stray values out. The hope is, then, that we can dismiss this talk of accepted statements that may nevertheless be false. Unfortunately, while it is true that you can reduce the occurrence of error with this approach, you can't eliminate it.

To see why this is so, consider another example: significance testing. Significance testing is a standard experimental practice found in sciences as disparate as psychology, chemistry, and medicine. In each science, experiments are designed to test a null 'no-effect' hypothesis, h_0, by choosing a region of rejection within a well-defined sample space of possible outcomes. If evidence lies in this region of rejection, then h_0 is rejected. The region is selected so that if the appropriate experimental justifications of randomness, independence and their kin hold, then there is only at most a small chance ϵ that given the supposition that h_0 is true, evidence falling in the rejection region will be collected. Another way to put it is to say that if h_0 is true, the probability of mistakenly rejecting it is less than the specified value of ϵ. Often ϵ may be made as small as one likes by increasing the sample size. Put in practice, we sample, check that the experimental assumptions hold, and then, should the sample obtained fall in the rejection region, we reject the null h_0 which states that the controlled variable has no effect. Note that the rejection of h_0 isn't hedged, but full-out; for instance, we *reject* the hypothesis that cigarette smoking has no effect on cancer rates in mice and men. The *grounds* for rejecting h_0 rest on the statistical –and *ipso facto* uncertain– claim that there is only a small, preassigned chance that we shall do so mistakenly.

Notice that what we've spelled out is similar to the structure of measurement

procedures. The proposal to buy certainty at the cost of taking additional measurements fails because our best experimental methods are themselves fallible. There are two points to notice about this result. First, error cannot be eliminated but only, even under the best of circumstances, controlled. Corrected measurements do not eliminate the possibility of accepting a false evidence statement. Patterns in evidence reports—how values may cluster around more than one point, the cardinality of data points in each of these neighborhoods, whether there are non-random trends in data through time—provides a wealth of information about whether the pattern of error fits what we would expect using a measurement procedure (or, more generally, an evidence gathering procedure) whose distribution of error is assumed to be normal. But these methods do not eliminate error, no matter how rigorous our methods. The tests and a well-designed s-default give us very good reason to be practically certain that each of our accepted evidence statements is true. Yet, at the same time, we may be practically certain that some relatively small proportion ϵ of our accepted evidence statements are plain false. Second, controlling error is expensive. We get the best (but not certain) results when we are keenly aware of what kind of errors we're liable to commit and design experiments or conduct measurements in a manner that reduces those risks. Not knowing all the ways one can go wrong contributes, in part, to the difficulty of identifying and correcting errors: we're ever discovering new ways to err. What can make this problem particularly difficult is deciding whether one has stumbled upon a new way of bungling, is dealing with a faulty instrument or an unusual sample, or is in the position of needing to reject some part of a well-confirmed theory.

It is worth pointing out that we are not promoting inconsistency tolerant logics for novelty's sake, nor are we making an ontological claim that the world itself is inconsistent, nor do we claim that mathematics would be much better off on naïve foundations. What we claim is that there are different sources of inconsistency when modeling operations that preserve properties like acceptance. One kind of inconsistency arises from using the logic to represent a collection of rules that yield a set of inconsistent formulae. Another kind of inconsistency arises when we represent within the theory itself a property of the inference scheme used within the logic, in this case a property of acceptance yielded by application of s-default rules. Ideally, we would like to do without either kind of inconsistency. But in practice it may only be practical to consider the first kind of inconsistency a target for correction, which is the subject of belief revision, and adopt a strategy of control for the latter. Living with the latter kind of inconsistency is part of the bargain of accepting defeasible conclusions. What statistical default logic does is to make this property explicit and control it with bounded-closure conditions.

4 Mechanizing Logic

It can be a complicated matter determining the membership of a statistical default extension, just as it can be a non-trivial matter determining the mem-

bership of default extensions for a default theory. Fortunately there has been considerable research in computational logic addressing this issue, both by exploring the inter-relationships between various non-monotonic formalisms and also by exploring the computability of this class of logics. A natural link between computability and default logics is *logic programming.*[13] While there are recent results establishing a correspondence between an important fragment of statistical default logic and logic programming [Wheeler and Damásio, forthcoming], we will focus here on more general correspondence results between semantics for logic programs and semantics for default theories. Again, our interest is to give an overview of the resources available and guide interested readers to references in the literature.

Logic programming arose from work begun by logicians in the 1950's with an aim to effect logical reasoning by computation. The first developments were automated theorem provers, which formed the theoretical basis for logical artificial intelligence[14]

What these papers introduced to computer science was the notion of *declarative*, as opposed to *procedural*, semantics. The idea underpinning declarative semantics is that a programmer should only concern himself with the declarative meaning of his programs while the procedural aspects of the program's execution are handled automatically. Logic programming [Colmerauer et. al. 1973; Kowalski 1974, 1979; Warren and Pereira 1977], or *Prolog*, became a privileged tool approximating this idea.

Work since has concentrated on development of a precise semantics for logic programs. Of one particular interest is the definition of negation within logic programs, since logic programs do not use classical boolean negation but rely instead on a non-monotonic operator, often called "negation by failure" or "negation by default". The non-monotonicity of this operator allows one to view logic programs as a special class of non-monotonic theories.

Indeed, one property that both epistemic relations and causal relations share that distinguishes both from logical implication is that the former pair are unidirectional in the sense that there is no implicit contraposition. This directionality of epistemic and causal relations is an essential feature of logic programs, 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.

[13] The relationship between logic programs and default theories was first explored in [Bidoit and Froidevaux 1988], where a stable model semantics was shown equivalent to a special case of Reiter default extensions, and has been the subject of subsequent work [Alferes *et. al.* 1995], [Alferes and Pereira 1996] [Alferes *et. al.* 1998], [Damásio *et. al.* 2001].

[14] Many of the foundational papers in AI are found in [Feigenbaum 1963].

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.

4.1 Logic program semantics

The semantics we will present for logic programs is the extended well-founded, *WFSX*, set forth in [Alferes and Pereira 1996]. Our interest here is not to provide a comprehensive account of the properties of this semantics, but rather present enough of it to establish the correspondence results between logic programs and default theories.

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 $T \cup \ not\ F$,[15] 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:

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

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 $\hat{I} = 1 - I(l)$

[15] Where $not\ \{a_1, ..., a_n, ...\}$ stands for $\{not\ a_1, ..., not\ a_n, ...\}$.

- 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

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

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 **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 **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 **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 4 ***(3-valued least model)*** *The 3-valued least model of a non-negative program is:*

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

Theorem 5 *$least(P)$ 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.

4.2 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^2_Δ 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 $F \in lfp(\mathsf{E}^2_\Delta)$;

- F is *false* in Δ w.r.t. the well-founded semantics if and only if $F \notin gfp(\mathsf{E}^2_\Delta)$;
- 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:

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

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

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 view 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.

4.3 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:

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

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

Theorem 7 *(Enforcedness) If E is an Ω-extension then E is enforced.*

Theorem 8 *(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 $E = E_1 \cup E_2$.*

4.4 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 9 *(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.

5 Epistemic Tools

To recap, what we've done is sketch a formal representation of a class of non-monotonic inference forms found at the heart of standard inferential statistics using a variation of default logic, called statistical default logic. We then briefly

discussed some semantics for default theories and logic programs, enough to give a sketch of how the correspondence results are obtained. While we think the example has intrinsic interest, we advanced it as being structurally analogous to an important class of relations that figure in contemporary theories of knowledge. We then presented a whirlwind tour of just some of the theoretical and computational resources available for modeling such relations, our aim being to introduce readers to this area of research, persuade them that there now exist enough theoretical infrastructure to support more precise definitions of epistemic relations, and demonstrate how one might proceed.[16]

What we find interesting about this proposal to call upon the resources of epistemology and artificial intelligence to study epistemic relations is the prospect of constructing *epistemic tools*, by which we mean specified relations. There are two areas where such tools can be of service. First, epistemic relations may—in so far as they can be represented within logic programming or some other computational logic framework—be tested. This is a non-trivial point for epistemologists, since we know from the foundations of mathematics that principles that appear obviously true to our intuitions may simply not be satisfiable. If we have a good understanding of the relations of our theory and are able to separate their behavior from the stated properties of epistemic notions, this would count as a significant advance by offering theorists the advantage of pinning down which parts of their theories to revise.

Another sense of epistemic tool arises in the event of successfully encoding an epistemic relation. *Knowledge representation and reasoning* is at heart an optimization problem, one that wishes to maximize the expressiveness of the representational language yet also maximize the power of the inference operations. Hence, the more we learn about actual epistemic relations the better position we all will be in to judge what is optimal.

Acknowledgement 10 *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.

[16] We should note that the scope of interdisciplinary coöperation between epistemology and logical artificial intelligence may be broader than what we've identified here as the core, including topics of agency, action, revision, explanation, multi-agent argumentation, and game-strategy.

[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] Bonjour, L. and E. Sosa. 2003. *Epistemic Justification*, Oxford: Blackwell Publishing.

[7] Carnap, R. 1950. *The Logical Foundations of Probability*, Chicago: Chicago University Press.

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

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

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

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

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

[13] 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.

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

[15] Elio, R. [ed.] 2002. *Common Sense, Reasoning, and Rationality*, Oxford: Oxford University Press.

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

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

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

[19] Foley, R. 1987. *The Theory of Epistemic Rationality*, Cambridge, Mass: Harvard University Press.

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

[21] 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.

[22] 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.

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

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

[25] 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.

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

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

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

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

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

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

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

[33] Kyburg, H. E., Jr. 1987. "Bayesian and Non-Bayesian Evidential Updating", *Artificial Intelligence*, 31:271-294.

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

[35] Kyburg, H. E., Jr. and C. M. Teng. 1999. "Statistical Inference as Default Logic", *International Journal of Pattern Recognition and Artificial Intelligence* 13(2) : 267-283.

[36] Kyburg, H. E., Jr. 2001. "Real Logic is Nonmonotonic", *Mind and Machines*, 11(4): 577-595.

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

[38] Mayo, D. 1996. *Error and the Growth of Knowledge*, Chicago: University of Chicago Press.

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

[40] 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.

[41] Pollock, J. 1986. "The Paradox of the Preface", *Philosophy of Science*, 53: 246-58.

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

[43] 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.

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

[45] Reiter, R. 1978. "On Closed World Data Bases", in Gallaire, H. and J. Minker, [eds.] *Logic and Data Bases*, 55-76.

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

[47] Rott, H. 2001. *Change, Choice and Inference: A study of belief revision and non-monotonic reasoning*, Oxford: Clarendon Press.

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

[49] Uzquiano, G. 2004. "The Paradox of the Knower without Epistemic Closure?", *Mind*, 113(449): 95-107.

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

[51] 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.

[52] 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.

[53] Wheeler, G. R. and C. Damásio. "An implementation of statistical default logic", forthcoming.

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

[55] Wheeler, G. R. 2004. "A resource bounded default logic", to appear in *The Proceedings of the 10th International Workshop on Non-monotonic Reasoning (NMR-2004)*, Whistler Village, British Columbia, June 2004.

[56] Weinburg, J. "Can one challenge intuitions without risking skepticism?", unpublished manuscript, Department of Philosophy, Indiana University.