# Statistical Defaults and Paraconsistency*

Gregory R. Wheeler
Center for Research in Artificial Intelligence (CENTRIA)
Department of Informatics, New University of Lisbon
2825-114 Caparica, Portugal
greg@centria.fct.unl.pt

December 19, 2002

**Abstract**

This paper discusses the relationship between the property of paraconsistency and *Statistical Default Logic*, a nonmonotonic logic that models common inference forms found in classical statistical inference such as hypothesis testing and the estimation of a population's mean, variance and proportions. Statistical Default Logic is paraconsistent in the sense that a set of default inference forms may induce an inconsistent set of conclusions yet not, by that fact alone, trivialize the statistical default consequence relation. The paper also clarifies a distinction between the studies of implication, reasoning and inference. This distinction is useful for understanding how to evaluate statistical defaults *vis a vis* probabilistic approaches to statistical argumentation and, it is suggested, useful for evaluating non-standard logics.

**Keywords:** Non-monotonic logic; statistical reasoning; resource-bounded inference; paraconsistency; Dutch Book.

## 1 Implication and Reasoning

In this paper I discuss the relationship between paraconsistency and *Statistical Default Logic*, a formalism that captures key structural features of statistical inference and allows for the logical representation of statistical arguments [Wheeler, forthcoming]. However, before turning to this discussion it is important to clarify precisely what kind of problem Statistical Default Logic addresses. I propose to do this by way of clarifying an ambiguity in the term 'logic', one that has at once spurred the development of non-classical logics while frustrating our attempts to evaluate their bounty.

The term 'logic' may be used to denote three distinct sorts of theories. In its most widely accepted and contemporary sense, 'logic' refers to the theory of

*This research was supported in part by the Institute for Human and Machine Cognition, Pensacola, and National Science Foundation grant SES 990-6128.

implication.[1] The theory of implication is a branch of mathematics that, like other branches of mathematics, consists of examining sets on which relations and functions have been defined and then determining what relations hold among these structures. It is a descriptive theory.

Normative questions arise when we consider which class of structures is appropriate for modeling some or other concrete problem. For the theory of implication, the problem is found at the heart of mathematics itself: the relation of logical consequence. The relation, ranging over propositions, holds in virture of the *logical form* of propositions. The Bolzano-Tarski formulation of logical consequence is the one most familiar to us whereby a proposition $\phi$ is a logical consequence of a set of sentences $\Gamma$ if in any model in which every sentence in $\Gamma$ is true is also a model in which $\phi$ is true. We should keep in mind that Tarski proposed his formulation of logical consequence as a mathematical tool, one whose purpose was to model our pre-theoretical notion of logical consequence [Tarski 1930, p 414]. Unlike other branches of mathematics whose problems often provide guidance in model selection in the form of physical constraints or observational evidence, the question of which class of structures is appropriate for modeling logical consequence is a philosophical one whose range of allowable answers is constrained most sharply by considerations from within the foundations and practice of mathematics.

The theory of implication must be distinguished from a second sense of 'logic', one that refers to the theory of reasoning. The theory of reasoning, if one were had in full, would be an applied mathematical discipline, one that studies a class of rational activities performed by humans. The theory of reasoning is far less developed than the theory of implication in part because the theory of reasoning depends upon rationality and rationality is not well understood [Harmon 2002]. Nevertheless, fragments of a theory of reasoning are found in the fields of (mathematical) psychology, economics, artificial intelligence, philosophy and the cognitive sciences. Because reasoning is bound to the notion of rationality, the theory of reasoning is, at least in part, normative. But being a human activity the theory of reasoning is, at least in part, descriptive. Sorting out precisely how each of these components constrains the other is another fundamental problem facing a theory of reasoning.

But it should be noted that the theory of implication is a poor choice for serving as either a descriptive or normative theory of reasoning. It fails as a descriptive theory of reasoning on at least three fronts. First, the theory of reasoning at best concerns operations performed on propositional attitudes, not propositions. But how should we represent a particular agent's belief states? One proposal is to give a probabilistic interpretation of belief and treat an agent's change in belief in a proposition as following a set of axioms. I'll have more to say about this proposal soon, but notice that as a descriptive theory it is implausible: I for one have not the slightest idea how to identify the supposed real number value of anything of mine that I would identify as a belief. Another

[1] There are two senses of the term 'implication', one having to do with the relation between antecedent and consequent in a true conditional, the other sense having to do with a relation between a set of propositions and a single proposition. The concern here is this latter sense.

proposal is to represent the propositional attitudes of belief and knowledge as operators on propositions. But it isn't clear what properties are sufficient for belief and knowledge, and the minimal set of conditions that philosophers have considered necessary for knowledge court paradox [of knowledge: Montague 1970; Cross 2000; Uzquiano, forthcoming; of knowability: Fitch 1963; Edgington 1985; Williamson 2000; Wansing, this volume.].

The second reason to doubt that the theory of implication describes human reasoning is that empirical evidence suggests that people generally aren't that good at deduction, which may explain why it is considered a skill. If a theory of implication were a good candidate for a theory of reasoning we should expect deductive derivations to be among the easiest tasks for us to perform. But we often do not see the logical consequences of a set of propositions; indeed, this situation describes the state of significant and open problems in mathematics. Furthermore, test subjects repeatably perform poorly, even if better than chance, when they are called upon to correctly classify simple candidate deductive derivations as valid or invalid [Evans, Newstead, and Byrne 1993].

The third reason against proposing the theory of implication as the descriptive theory of reasoning is the resource constraints each, the subject and the logic, make on the other. First, human beings have neither an infinite memory nor the luxury of eternity to work out all logical consequences of their beliefs. That we run out of cognitive capabilities long before applicable rules is also a problem for a normative theory of reasoning, a point we will return to shortly. For now it is important to notice that human reasoning imposes its own set of awkward constraints on the theory of implication, for reasoning occurs within time and a situation, very often involving less than certainly known but rationally accepted premises and often proceeding in a non-monotonic fashion. But rational acceptance doesn't behave like truth in a model, logically derived consequences are anything but defeasible, and situations and time are neither among the basic elements of a theory of implication nor are they features easily added without weakening the theory itself.[2] Finally, without a consistent set of propositions or marginal probability assignments the theory of implication and its probabilistic cousin trivialize. But people don't necessarily stop reasoning well when their beliefs are inconsistent. For one thing, they may not be aware that their beliefs are inconsistent. For another, casual observation notes that people typically don't suffer fits of gullibility when they are made aware of having inconsistent beliefs.

The failure of the theory of implication to serve as a remotely plausible candidate for a descriptive theory of reasoning suggests that people have in mind proposing the theory of implication as a *normative* theory of reasoning, perhaps as a description of an ideally rational agent. For example, Bayesianism offers in one package the normative notion of Bayesian rationality along with a method for changing degrees of belief for an ideal Bayesian agent. Likewise, symbolic

[2]Artificial Intelligence theorists are well aware of the tension between a formal language expressive enough to suitably model the logical form of natural language expressions yet powerful enough to support inference. For discussion and an example of how to juggle this tension, see [Schubert and Hwang 2000].

approaches offer prescriptive advice for how certain propositional attitudes, such as belief and knowledge, behave when we accept certain constraints on those concepts, delivering to us a guide to what an ideal agent would conclude given those constraints.

While the study of ideal reasoning is fruitful for as far as it goes, it is not clear precisely what prescriptive advice we should take away for the theory of human reasoning. Fitch's paradox remains for modal approaches, and formulating a plausible, paradox-free epistemic closure principle remains an open problem for a language with a knowledge predicate.[3] On the other hand, proponents of Bayesian rationality argue that our degrees of belief (!) should be isomorphic to the probability measure since otherwise we may with certainty lose money if presented with a 'Dutch Book', which is a set of bets offered that returns a certain loss no matter the outcome. But it is fair to ask how well this model accords to the range of things we are able to do: Are we psychologically able to approximate the conditions that bound ideal agents? To answer this question, we must address whether one is able to assign a real-numbered value to a belief he may have, whether he may compare all of his beliefs and, furthermore, whether he may do so consistently in order to thwart clever bookies. The reason that it is important to consider the sense in which these operations are possible is that this model of ideal reasoning is unforgiving to those of us who fall short. If we are prevented from meeting these conditions by psychological necessity, say, we may then ask what force the prescriptive advice from such a theory holds *for us*.

The point I wish to stress is that the relationship between the reasoning we perform in our heads and the algebras we study is not at all well understood. This point brings us to consider the third sense of the term 'logic', the one that figures in the remainder of this paper and is sometimes called the study of entailment. For reasons to be considered next I call this sense of 'logic' the *theory of inference*.

## 2 Entailment and The Theory of Inference

The term 'entailment' is commonly understood to refer to a relation that sometimes holds between the premises and conclusion of an argument. Traditionally, the term is used as another name for the sense of implication we have been working with here, the relation designed to model logical consequence. This convention of treating 'implication' and 'entailment' as working synonyms suits

[3]Note on epistemic closure: Setting aside the paradoxes for the moment, notice that simply closing knowledge (of some agent S) under implication is not a suitable principle. The principle 'If S knows p and p implies q, then S knows q' fails to specify that S knows q by reason of deriving it from p. Yet building this qualification into a suitable closure principle, as in 'If S knows p and S believes q on the basis of recognizing that p implies q, then S knows q' forms the basis of work on the problem of epistemic closure, a project that at once seeks to avoid counter examples and specify precisely what it is for a person to recognize an implication of what one knows. It should be noted that some [e.g., Dretske 1970 and Nozick 1981] have rejected the claim that knowledge is closed under implication.

when we restrict ourselves to the study of logical consequence. But if we understand at least one sense of 'logic' to refer to the study of the structure of *arguments*, then the notions of entailment and implication come apart.

It is here that we must be careful. Framing logic as the study of entailment is freighted with a controversial history, for the claim that logic is the study of entailment has served those who would have classical logic replaced by some markedly different system [e.g., Anderson, *et. al.* 1975, 1992; Routley *et. al.*, 1982; Priest 1998]. We've already suggested that figuring out which class of structures is the right one for the theory of implication is a philosophical matter of picking the appropriate class of structures to model the working notion of logical consequence used within mathematics itself.[4] It is my view that the development of the theory of implication should not to be guided by events in physics, nor from observing how we reason out fictional plots involving impossible objects and certainly not from our apparent day-to-day ability to reason with inconsistent beliefs. The issues these problems raise are problems of applying a logic to matters entirely distinct from understanding logical consequence.

Furthermore, we should be careful to distinguish between the psychological act of drawing a conclusion from an argument we're considering and the abstract thing that is an argument, the latter being constructed from sentences and having relations that all parties involved may describe as holding (or not) among some sentences under consideration. The latter is a thing liable to modeling by defining relations and functions on sets of propositions in ways familiar to us while the former, as we've seen, is not transparently so.

With these distinctions in mind, it can be seen that there is room for more than one entailment relation to tie a conclusion to its premises. The distinction between deductive and inductive inference suggests itself here. But whereas the theory of implication has served as a good template for understanding a certain class of deductive arguments, perhaps all of them, our failure to understand the structure of inductive arguments has been widely noted and, in some quarters at least, definitive. But the task of developing an inductive logic has suffered both from a poorly organized catalog of good inductive argument patterns and a popular zeal for a particular *monotonic* tool: probability theory.

What the study of induction calls for is a reasonably constrained class of arguments with which to work; only then may we have a better idea of what parameters to fix in order to evaluate proposals. Classical statistical inference offers a good class of candidates to the inductive logician in large measure because the structure of both statistical method and theory lends itself to abstraction. If classical statistical inference indeed provides a suitably constrained class of argument patterns, then we may consider precisely what kind of mathematical tool is best suited to modeling them.

[4]The view that the aim of logic is to find the right class of structures is to be contrasted with *pluralism*, the view that there isn't a single correct notion of logical consequent but many [Restall and Beall 2000], and the view that there isn't a fixed answer to the question of which notion of logical consequence is the correct one because the logical vocabulary of a language is only fixed by convention [Varzi 1999].

## 3 Statistical Inference and Statistical Defaults

The entailment relation that figures in classical statistical inferences is most certainly not logical consequence. One property that distinguishes this entailment relation from logical consequence is monotonicity: classical statistical inference is non-monotonic, for reasons we will soon consider, while logical consequence is monotonic. Another property that distinguishes statistical inference from logical consequence is paraconsistency: a set of accepted statistical premises are defeasible, rather than true, and so may be inconsistent. Statistical consequence should not trivialize under these conditions. That classical statistical inference is in some sense tolerant of inconsistency sets statistical inference apart from probabilistic approaches to modeling uncertain reasoning. We will return to this point as well. But let us now consider what marks an inference as a classical statistical inference.

An important feature of classical statistical inference is its emphasis on the control of error. In making statistical inferences—a term intended to include hypothesis testing and estimating basic parameters of populations, such as their means, proportions and variances—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.

One type of classical statistical inference is significance testing. Significance tests are designed to yield a false rejection of the null hypothesis $H_o$ no more than a fraction $\hat{\alpha}$ of the time. A significance test is a good one if the test succeeds in controlling error—that is, if the frequency of error is in fact $\hat{\alpha}$—and the frequency of error is relatively small, usually less than 0.05.

Ronald Fisher's [Fisher 1956] canonical example suits our purpose here. Fisher imagines a lady who claims to distinguish by taste the order of ingredients, milk first or tea, added to her cup. Question: How do we test whether she can do it? The short answer is to see if she performs better than chance. Performing *no* better than chance, Fisher reasoned, would amount to correct calls of heads on a sequence of fair coin tosses. If the null claim $H_o$ is true then we would expect that the probability of correctly identifying the ingredient-order of, say, a 5 cup series of milk-tea mixtures to be 1 out of 32, or 0.031. If we accept these odds as long enough, then we have the rudimentary conditions for rejecting the null and inferring that she has the ability: see if she can correctly classify a series of 5 prepared cups of tea.

But there is a catch. Five correct guesses is significant only if the sequence of her correct guesses is not had by means other than her sense of taste. If the milk-first samples differed in color from tea-first samples, for instance, her scoring 5 out of 5 wouldn't tell us much about the veracity of her claim. This significance test is bounded in error by $\hat{\alpha} = 0.031$ when in fact the chance we face of rejecting the null and yet her not having the ability to discriminate milk-first from milk-last cups of tea-milk mixtures is just 1 in 32 [Wheeler, 2000].

Ideally, we want the sample—her sips and classification of the five cups of tea—to be representative of the target population, that is the entire class of her sips and classification of cups of tea and milk. Textbooks tell us that to achieve representative samples we need to ensure that the sample is drawn at random [Cramér, 1951; Moore, 1979; Baird, 1992]. But demanding that a selected sample be selected by a method that yields each possible sample with equal long-run frequency is difficult if not impossible to meet. In practice, the grounds for accepting that a sample is representative rests on failing to detect that it is biased. That is, given a sample and the *absence of specific information* that would indicate that the sample is biased, we infer that the sample is representative of the parameter(s) of interest in the target population. So given that a lady correctly identifies 5 of 5 samples and it isn't the case that we know of conditions that would suggest a bias (e.g., it isn't the case that we know that the cups weren't stirred, nor is it known that the experiment wasn't performed double blind, nor known that the cups weren't standardized in size and weight) we reject the null and infer that her claim is true: she can do it.

It turns out that the structure of this inference pattern is standard across a significant class of techniques common in classical inferential statistics. We draw a sample, test for bias, then conclude (by default) that the statistical model 'fits' and thereby the values observed or measured in the sample hold in the population within the bound of error prescribed by the statistical model. The conclusion is made by default since new information may come to light and incorporated into the theory that signals that the sample may be biased and so the inference should be withdrawn. It is important to lay stress on the kind of evidence that triggers a withdrawal of a statistical conclusions. It is often the case that we don't have outright evidence of a sample being biased. Rather, we learn of conditions that undermine our confidence that the sample is representative: recording 5 of 5 correct guesses of unstirred cups of tea doesn't entail that the lady's record of correct guesses isn't due to her ability. The conclusion to draw is that the test is not a good one and so no conclusion about the lady's ability should be based on *this* inference.

The non-monotonic behavior suggested by this description is common in statistical reasoning. What is interesting is that the logical structure of this inference pattern is very similar to default rules found in default logic. Although a connection between statistical inference and default logic has been suggested in the work of [Tan 1997], the first proposal to represent classical statistical inference in terms of defaults is [Kyburg and Teng, 1999]. A default is an inference rule of the form

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

interpreted roughly to mean that given $\alpha$ and the absence of any negated $\beta_i$'s, conclude $\gamma$ by default [Reiter, 1980]. Kyburg and Teng observe that default rule justifications—the formulae denoted by the $\beta_i$'s in the default inference form—represent the role randomization is thought to play as a sufficient condition for a sample being representative.

We've already observed that randomization isn't a necessary condition for drawing a representative sample: it is often impractical or impossible to draw a random sample. A more important consideration, however, is Kyburg and Teng's claim that randomization is not a sufficient condition, either. It is not sufficient since a randomly selected sample may produce a sample that we know, given our background knowledge, is not representative. For example, a random sample of shelled walnuts from a barrel may draw only members from the bottom of the barrel. But if our interest is to determine the proportion of broken to whole walnuts, we know that this sample is not representative even if drawn at random: knowing that broken bits settle to the bottom of the barrel trumps knowing that the sample is a random one.

We began the discussion of statistical inference by citing the importance of controlling error. But notice that standard defaults do not accommodate this property, for there is no constraint in the logic to distinguish between the case when a battery of tests (default justifications) are suitable for the statistical model holding and the case when no tests for bias are even included at all [Wheeler, forthcoming]. Hence, standard defaults only provide half of the structure of statistical inference. A remedy is proposed in the notion of a statistical default. S-defaults differ from defaults by explicitly acknowledging the *upper limit* of the s-default's frequency of error.[5] Call a default in the form of

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

an *$\epsilon$-bounded statistical default* and the upper limit on the frequency of error-parameter $\epsilon$ an *$\epsilon$-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 $\alpha$ and no negated $\beta_i$'s, $\gamma$ is false no more than $\epsilon$ over the long-run application of that rule. (A Reiter default is a special case of a statistical default, namely when $\epsilon = 0$). A statistical default is sound just when the upper limit of the frequency of error is *in fact* $\epsilon$. An s-default is a good inference rule if it is sound and $\epsilon$ is relatively small, typically less than 0.05.

The art of constructing a good statistical default, one whose advertised $\epsilon$-bound is an acceptable level and in fact true, depends upon the background knowledge necessary to form the set of s-default justifications that determine the fit between a statistical model and its successful application. Disagreement about the soundness of a statistical default is then similar to disagreement about the soundness of a deductive inference in the sense that a distinction is made between the form of the inference and whether the constituents of the inference are (or not known to be) satisfied. Treating $\epsilon$ as an explicit parameter in statistical default forms allows there to be an explicit constituent in the argument form that, much like a premise in a deductive argument, may be true or false.

To illustrate, return to Fisher's test of the tea-taster's claim. Significance tests represented within a statistical default framework are inference rules de-

[5] A trivial corollary of the frequency of error $\hat{\alpha}$ for a statistical inference is the upper limit of the frequency of error, denoted by $\epsilon$. So, if $\hat{\alpha} = 0.03$ is understood to mean that the probability of committing a Type I error is 0.03, then $\epsilon = 0.03$ is understood to mean that the probability of committing a Type I error is no more than 0.03.

signed to yield the conclusion $\gamma$ when $\gamma$ is false no more than $\epsilon$ over the long run application of that rule. The frequency of error of a five-trial binomial-chance model is 1/32, which is a mathematical truth. Whether the frequency of error associated with this statistical model represents the risk of error we in fact face in rejecting $H_o$ after seeing 5 correct guesses, however, is an empirical matter: it depends on how thorough a job we do in checking for bias, which is the substance of the statistical default justifications. Consider these substitutions for schema (2):

> $\alpha$ : A sample of 5 milk-tea mixtures were prepared and the lady correctly identified the mixture order in all 5.
>
> $\gamma : \neg H_o$ : the lady does have the ability to discriminate order by taste.
>
> $\beta_1$ : The sample on which this inference is based is unbiased.[6]
>
> $\beta_2$ : The sample of 5 is the largest appropriate sample we have.
>
> $\beta_3$ : There is no known set of prior distributions such that conditioning on them with the data obtained leads to a probability interval conflicting with $\epsilon$.[7]
>
> $\epsilon = 0.03125$

We accept that the subject has the ability to discriminate by taste (reject the null claim that she doesn't) only if no s-default justifications are violated.

For instance, violating justification $\beta_1$ amounts to knowing two things. First, we know that the sample comes from an unusual class of samples and, second, that the proportion of samples that lead to the false rejection of $H_o$ in this class in fact leads to falsely rejecting $H_o$ more often than 3% of the time.

The second justification demands that we not know that there is no larger sample. It is possible that there is data that we are unaware of: $\beta_2$ doesn't demand that $n$ is the largest sample we have, but that we don't know that it isn't.

Finally, the third justification pertains to prior knowledge. If we knew that the relative frequency of error for $H_o$ was 0.25, as opposed to the $\alpha$ value of 0.03, that would surely have a bearing on our significance test inference. If we have prior knowledge of distribution for a parameter we should use that knowledge.[8]

[6] Although we cannot demand lack of bias as a prerequisite, we could demand such prerequisites as "good sampling technique." Nevertheless, since we cannot calculate the added frequency of correct conclusions due to good technique, this seems a questionable prerequisite.

[7] What "conflict" comes to is an interval that overlaps with, but does not include and is not included in, $[1-\epsilon, 1]$.

[8] See [Kyburg and Teng, 1999] for examples of statistical estimations represented as default inferences, and [Wheeler, 2002] for s-default corollaries and, in addition, an s-default example representing conditionalizing with objective probabilities.

# 4 Inconsistent Exenstions and Dutch Books

A statistical default theory is analogous to a default theory[9] in that statistical defaults appear in the object language and an s-default theory induces non-monotonic consequences via a fix-point construction. A non-monotonic consequence relation is then defined by identifying the conclusion set with the intersection of all candidate extensions.

Given the parameter $\epsilon$, however, a complication arises in building suitable extensions. The problem centers on the need to restrict membership of any conclusion set to just those formulae appearing in an extension that may be reached by a sequence of inference steps that remains under a specified error bound. But bounds for frequency of error for particular statistical inferences do not necessarily carry over unchanged when chained together to form a statistical argument. Accepting a hypothesis $H_1$ with a confidence 0.95 and accepting another hypothesis $H_2$ independently with a confidence 0.95 doesn't entail the acceptance of $H_1$ and $H_2$ at 0.95. The upshot of this observation is that closure conditions within a statistical default theory must be relativized to a given error bound. Since the point of an $\epsilon$-bound is to provide a firm upper-bound on frequency of error, an error bound for a sequence of inference steps is calculated by summing the error bound parameters of all constituent inference steps.[10]

It should be noted that summing error bounds is an imprecise measure: it assumes the worst case, that the frequency of error for statistical inference is independent. However, this result accords to practice: reduction of false positives, or Type II errors, depends on background knowledge in a manner that Type I errors do not. If $\phi$ appears in all extensions yet cannot be reached within the designated error bound, then there is an incentive to refine the statistical inferences within the theory in an attempt to accommodate $\phi$. One way to do that is to look for dependency relations between statements that would allow their conjunction to have an error bound less than the sum of the conjuncts; another would be to revise the statistical tests applied in a sequence of reasoning by lowering their error bounds. Regardless, a feature of SDL is that by having the error-bound parameter function independently of the mechanism for inducing consequences, one may isolate formulae that are non-monotonic consequences of a theory but not within current bounds. This feature allows one ot focus on what part of the theory to first consider revising.

A candidate statistical default extension is constructed much like a candidate default extension. Recall that a default extension on a default theory $\langle W, D \rangle$ is built sequentially by first closing the set of sentences, $W$, under logical consequence, applying all applicable defaults in $D$ to the set of consequences of $W$, closing that set (extension) under consequence, and so on. While a standard default extension is built sequentially by alternatingly closing an extension under consequence and applying defaults until no more defaults can be applied, statistical defaults are built in the same manner but only with those sentences

[9] 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 formulae.

[10] For details, see [Wheeler, forthcoming].

that are below the bound set by $\epsilon$.

To illustrate how a statistical default extension is constructed, consider the following example.

**Example 4.1**. Let $\Delta_s^1 = \langle W, S\rangle$ be a statistical default theory, where $W =$ and $S$ contains four s-defaults:

$$D_s = \{\frac{: A}{A}0.01, \frac{: B}{B}0.01, \frac{A : B, C}{C}0.01, \frac{A \wedge B : \neg C}{\neg C}0.01\}$$

For an error-bound parameter $\epsilon_1 = 0.02$, there is one statistical default extension $\Pi^1$ containing

$$A, B, A \wedge B, C.$$

The bounded sentence $A$ at $\epsilon_A$ is included in extension $\Pi^1$ by applying the default $\frac{:A}{A}$ and bounded sentence $B$ at $\epsilon_B$ is included by applying the default $\frac{:B}{B}$, where each inference has an error bound of 0.01, so $(A)_{0.01}$ and $(B)_{0.01}$. $(A \wedge B)_{\epsilon_{A\wedge B}}$ is included in the extension, since the sum of the error bounds of conjoining $A$ and $B$ is 0.02, that is $(A \wedge B)_{0.02}$. The bounded sentence $C$ at $\epsilon_C$ is included by using $A$, whose error bound is 0.01, to apply the default $\frac{A:B,C}{C}$, whose error bound is also 0.01. Hence $(C)_{0.02}$. The default $\frac{A\wedge B:\neg C}{\neg C}$ cannot be applied because the resulting conclusion $\neg C$ would have an error bound of 0.03, $(\neg C)_{0.03}$ which is above the designated threshold $\epsilon_1 = 0.02$.

For an error parameter $\epsilon_2 = 0.03$, there are two statistical default extensions $\Pi^1$ , which is the same as described above, and $\Pi^2$, where $\Pi^2$ contains

$$A, B, A \wedge B, \neg C.$$

The default rule that could not be applied before is now applicable with respect to $\epsilon_2$, giving rise to the second extension $\Pi^2$.[11]

Like their default logic counterparts, it is not uncommon for statistical default theories to have multiple extensions. We see in Example 4.1 that $\Delta_s^1 = \langle W, S\rangle$ when $\epsilon = 0.03$ is inconsistent with respect to the pair of applicable s-defaults concerning $C$.

However, we've yet to define a consequence relation. A common way of defining a non-monotonic consequence relation for a default theories is to include only those sentences in the conclusion set that appear in every extension.

[11]The complete extensions $\Pi^1$, when $\epsilon = 0.02$, $\Pi^1$ and $\Pi^2$, when $\epsilon = 0.03$, are as follows: $\Pi^1_{\epsilon=0.02} = \{A, B, A \wedge B, C\}$; $\Pi^1_{\epsilon=0.03} = \{A, B, A \wedge B, C, A \wedge C, B \wedge C\}$; $\Pi^2_{\epsilon=0.03} = \{A, B, A \wedge B, \neg C\}$.

An analogous consequence relation for statistical default theories is defined as follows:

**Definition 1.** *Skeptical Statistical Consequence*: Let $\Delta_s = \langle W, S \rangle$ be a statistical default theory at $\epsilon$, $A$ a sentence and $\Pi$ an extension on $\Delta$ bounded by $\epsilon$. Then $A$ is a skeptical consequence of $\Delta_s$ at $\epsilon$—written, $\Delta \mid\!\sim_\epsilon A$—just in case $A \in \Pi$ for each extension $\Pi$ on $\Delta_s$ at $\epsilon$.

The consequence relation $\mid\!\sim_\epsilon$ is nonmonotonic. Notice that by either augmenting the set of (bounded) sentences in the $W$-component of a statistical default theory or adding new default rules to the $S$-component a previously induced statistical consequence (at a particular error-bound) may then fail to remain supported by the statistical default theory (at that particular error-bound). Citing results from [Wheeler, forthcoming], we remark also that s-default logic is supra classical (i.e., when the set of s-defaults is empty, the consequence relation is logical consequence) and that Reiter default logic is a limitting case of S-default logic, namely when every s-default in an s-default theory has an error-bound of 0).

Finally, notice that statistical default consequence is paraconsistent. The $S$-component of a statistical default theory may contain s-defaults that induce consequents that are mutually inconsistent. But, since statistical consequence is a skeptical inference operation, only those formulae that appear in every $\epsilon$-bounded extension will be $\epsilon$-bounded consequences of the theory. It is important to stress that this paraconsistent behavior is common to default logic. The novelty lies in how we may think about building a statistical theory. While it is of course preferable to avoid introducing an inconsistency into a statistical theory, it is not catestrophic to do so—when in the form of adding s-default rules that induce an inconsistent set of consequents, that is. We may have good grounds for including each s-default a set that is nevertheless inconsistent; our working view of a theory is rarely from a bird's eye. Furthermore, a logic that works in this circumstance provides the rudiments for tools for isolating the portion of the theory to revise. This result is in sharp contrast to probabilistic approaches.

The point to draw from the contrast is this: We shouldn't be cowed by Dutch Book arguments made to the effect that the machinery of s-default logic is, in so far as it allows us to encode a statistical theory that is not stand one-to-one with the probability measure, irrational. It bears repeating that s-default logic is designed for sentences that a community is in a position to accept and not a tool for representing beliefs in an imaginary agent's head. Furthermore, accepting a framework that accommodates an inconsistent theory is not to endorse having an inconsistent theory: on the contrary, it gives us some resources for making a revision. We remarked in sections 1 and 2 that Bayesianism doesn't offer a *prima facie* attractive framework because of the implausible assumptions it demands. Nevertheless, one may be persuaded by claims that Bayesian approaches represent 'the only game in town.' But now that we've been introduced to SDL, we may see that the argument to the effect that rationality (inference) is lost when an accepted theory is not isomorphic with the probability measure is unsound.

## 5 Conclusion

In this paper I've argued that we must distinguish between at least three senses of 'logic' and that of those three the most fruitful one to keep in mind when evaluating a new logic is the study of entailment relations appearing in arguments. An informal introduction to statistical default logic was provided. This is a non-monotonic framework for representing arguments composed, at least in part, of classical statistical inferences. A non-monotonic consequence relation for statistical default logic was defined in the familiar skeptical fashion. This consequence relation allows for a mild form of paraconsistency, namely when the set of s-defaults induces a set of consequents that is inconsistent. Skeptical statistical default consequence does not necessarily trivialize for precisely the same reasons skeptical default consequence doesn't: there may be consequents that appear in all extensions, which represents the non-monotonic consequence set of the consistent fragment of the statistical default theory (at a particular error bound). While this is a common feature to skeptical consequence relations in general, the novelty rests in application: Statistical Default Logic offers a genuinely non-monotonic and mildly paraconsistent alternative to probabilistic accounts of classical statistical inference.

## References


[1] Anderson, A. R., N. Belnap and J. M. Dunn. 1975. *Entailment: The Logic of Relevance and Necessity, Vol. 1.* Princeton: Princeton University Press.

[2] Anderson, A. R., N. Belnap and J. M. Dunn. 1992. *Entailment: The Logic of Relevance and Necessity, Vol. 2.* Princeton: Princeton University Press.

[3] Baird, D. 1992. *Inductive Logic: Probability and Statistics*, Englewood Cliffs, New Jersey: Prentice-Hall Press.

[4] Cramér, H. 1951. *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

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

[6] Edgington, D. 1985. "The Paradox of Knowability," *Mind* 94: 557-568.

[7] Evans, J., S. E. Newstead and R. M. Byrne. 1993. *Human Reasoning: The Psychology of Deduction.* Hove, UK: Lawrence Erlbaum Associates.

[8] Fisher, R. 1956. *Statistical Methods and Scientific Inference.* Edinburgh: Oliver and Boyd.

[9] Fitch, F. 1963. "A Logical Analysis of Some Value Concepts," *The Journal of Symbolic Logic* 28: 135-42.

[10] Harmon, G. 2002. "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.

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

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

[13] Kyburg, H. E., Jr. 1996. "The Rule of Adjunction and Rational Inference," *Journal of Philosophy*, xx:xxx-xxx.

[14] 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.*

[15] Mayo, D. 1996b. *Error Statistics and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

[16] Moore, 1979. *Statistics*, San Francisco: W. H. Freeman Press.

[17] Neyman, J. 1950. *First Course in Probability and Statistics*. New York: Henry Holt and Co.

[18] Neyman, J. 1957. "Inductive Behavior' as a concept of philosophy of science," *Review of the International Statistical Institute*, 25:5-22.

[19] Priest, G. 1998. "What is so bad about contradictions?" *Journal of Philosophy*, 95(8): 410-426.

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

[21] Restall, G. and J. C. Beall. 2000. "Logical Pluralism," *Australian Journal of Philosophy*, 78: 475-493.

[22] Routley, R., R. Meyer, V. Plumwood, R. Brady. 1982. *Relevant Logics and their Rivals, vol. 1.* Atascadero, CA: Ridgeview.

[23] Schubert, L. K. and C. H. Hwang. 2000. "Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding," in L. Iwanska and S.C. Shapiro [eds.] *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, Menlo Park, CA: MIT/AAI Press, 111-174.

[24] Tan, Y. H. 1998. "Default Logic and Inductive-Statistical Reasoning," *Synthese* 110: 357-379.

[25] Tarski, A. 1936. "On the Concept of Logical Consequence," appearing in Alfred Tarski, *Logic, Semantics and Meta-Mathematics.* Princeton: Princeton University Press.

[26] Uzquiano, G. Forthcoming. “The Paradox of the Knower without Epistemic Closure?”

[27] Varzi, A. 1995. “Model-Theoretic Conventionalism,” appearing in *Karlovy Studies in Reference and Meaning*, Hill, J. and Koťátko

[28] Wheeler, G. R. 2000. “Error Statistics and Duhem’s Problem,” *Philosophy of Science* 67,3:410-420.

[29] Wheeler, G. R. 2002. “Kinds of Inconsistency,” appearing in *Paraconsistency,* Carnielli, W. et. al. [ed.], New York: Marcel Dekker.

[30] Wheeler, G. R. forthcoming. “Statistical Default Logic,” *JJJAA*, xxx-xxx.

[31] Williamson, T. 2000. *Knowledge and its Limits* Oxford: Oxford Press.