

# Inductive Learning and Defeasible Inference\*

Kevin B. Korb  
Dept. of Computer Science  
Monash University  
Clayton, Victoria 3168  
Australia  
korb@bruce.cs.monash.edu.au

**Abstract.** The symbolic approach to artificial intelligence research has dominated AI until recent times. It continues to dominate work in the areas of inference and reasoning in artificial systems. I argue, however, that non-quantitative methods are inherently insufficient for supporting inductive inference. In particular there are reasons to believe that purely deductive techniques—as advocated by the naive physics community—and their nonmonotonic progeny are insufficient for supplying means for the development of the autonomous intelligence that AI has as its primary goal. The lottery paradox points to fundamental difficulties for any such non-quantitative approach to AI. I suggest that a hybrid system employing both quantitative and non-quantitative modes of reasoning is the most promising avenue for developing an intelligence that can avoid both the paralysis induced by computational complexity and the inductive paralysis to which purely symbolic approaches succumb.

**Keywords.** Inductive inference, ampliative inference, defeasible reasoning, nonmonotonic reasoning, Bayesianism, machine learning, logicism, epistemology, lottery paradox, hybrid reasoning.

---

\* I wish to acknowledge the helpful comments of Noretta Koertge, Jon Barwise, Raymundo Morado, Chris Wallace and an anonymous referee on earlier drafts of this paper.

# 1 Learning as Inductive Inference

Alan Turing invented computer science in order to solve a problem posed by David Hilbert: Is first-order logic decidable? That is, given a sentence in first-order logic can we decide in finite time, and using only well-understood inferential steps, whether it is or is not a theorem? In order to solve this problem, Turing had to provide a more precise definition of decidability. He did so by providing a precise definition of a computation, giving us our concept of Turing machines. And he proved that first-order logic is not decidable (Turing 1937; to be sure, simultaneously with Church's proof of the same, 1936).

It is perhaps the existence of this connection at the foundations of computer science which has led many to believe that artificial intelligence is fundamentally an exercise in logic, that artificial inference is exhausted by deductive inference and so that artificial knowledge must be encoded in a logical language. There have been recent challenges to this paradigm from the neural computing community, based upon some limited successes in that approach. But it has not been well appreciated that this “logicist” paradigm for artificial intelligence is, and always was, confused.

We can obtain at least as many definitions of ‘artificial intelligence’ as we can find introductory textbooks. But surely we can say that what artificial intelligence research aims at is the production of a computer system which behaves in ways we deem intelligent; in particular, most researchers believe it is possible (in theory, eventually, given an indefinitely long wish-list of technological breakthroughs) to program a computer that could power a robot to behave as intelligently as humans do, much like Isaac Asimov's detective robot Daneel, say. Whatever intelligence may or may not be, that would satisfy us (and it would go well beyond satisfying Turing's original test for intelligence; cf. Harnad's “Total Turing Test” in 1989). Without getting into very deep points about such a test for intelligence, we can see that its adoption requires a computer system that can learn about the world into which it is born: there is no physically normal human who is unable to learn what foods are edible, what animals are dangerous, etc.—unless that human is *not* intelligent. Any system which cannot learn the simplest fact about the world is not intelligent. So whatever else intelligence involves, it requires the ability to *learn* about its world, which we can call empirical learning.

Such abilities are not well reflected in deductive logic: the distinguishing feature of a deductive inference rule is that it is valid. Validity means that whenever the premises of the inference are true, the conclusion is true—indeed, it is *guaranteed* to be true. The only way to guarantee the truth of a conclusion is to adopt a conclusion which, in some sense, tells us nothing over-and-above the premises we have already accepted; its content must already be implicit within the premises. Or: the possible worlds in which the premises are true are a subset of the possible worlds in which the conclusion is true. Such inferences are *non-ampliative*, they add nothing to our empirical knowledge.<sup>1</sup> While such deductive inference may supply us with learning of a sort (we “learn the implications” of what we already believed), it is not empirical learning. It is, therefore, not enough learning to pass our test of intelligence. The logicist research program in artificial intelligence, despite dominating much of the work in AI, has not yet succeeded in fulfilling its early promises. Indeed, *it can never succeed*, for it is trying to use non-ampliative inference to do what only ampliative inference can.

Ampliative inference is otherwise known as inductive inference: at its most general, any inference is inductive if it leads from a set of premises to a conclusion which *could* fail to be true even though the premises are true. There have been many varieties of inductive inference proposed by philosophers and scientists; some have turned out to be useful and others not. The simplest to describe is one of the least useful to us, enumerative induction.<sup>2</sup> In enumerative

induction we simply observe a sequence of objects and if the first  $n$  of them all have some property, we conclude that all of them have that property. For example, after observing many thousands of white swans, and swans of no other color, Europeans may be forgiven if they conclude that all swans are white (there are black swans in Australia). Clearly, the conclusion (if it were but true) tells us something new about the world that is not implicit in the premises (in this case, the observation reports).

The use of enumerative induction as an example may be misleading. A traditional way of distinguishing induction from deduction is to say that induction proceeds from particulars to generalizations and that deduction proceeds in the reverse direction. Enumerative induction is then used to illustrate the former. But such a definition is an oversimplification. There are deductive inferences that lead from particular to particular, and others that lead from generalization to generalization: any inference in propositional logic will provide an example, if you substitute particular (general) statements for its sentential variables. On the other hand, some kinds of inductive inference may lead from one particular to another; for example, from the current swan being white to the prediction that the next swan will be white (the inductive logician Rudolf Carnap was particularly fond of such cases of “instance confirmation”; cf. his appendix to his classic work 1962). It is not the “logical form” of an inference which determines whether or not it is inductive, but the informational relation between premises and conclusion. Pointing this out is not mere pedantry: it will be an important point later in the discussion that the default inference from particular to particular—for example, the inference from Tweety being a bird to Tweety being capable of flight—is a kind of induction.<sup>3</sup>

As inductive inference is defined by negation—those forms of inference that are not deductive—there are many kinds of inductive inference open to investigation; indeed, there are too many to investigate. But there is no disputing that *some* inductive processes work, for however it is that humans learn about their world, that process works and is inductive.

## 2 Overview

Bayesianism arose within epistemology as an account of rational belief and rational changes of belief in the work of de Finetti (1937) and Ramsey (1931). More recently, it has flourished as a theory of statistical inference (Lindley 1971), as a philosophy of science accounting for normative methodological standards for scientific research (Howson and Urbach 1989), and as a leading contender for accommodating reasoning under uncertainty within artificial intelligence (Pearl 1988 and Neapolitan 1990). What serves to distinguish all of these efforts as (loosely) Bayesian is the belief by their authors that the probability calculus provides the single best standard for assessing inference with incomplete evidence—as exemplified by the use of Bayes’ Theorem for conditioning upon experimental or observational evidence—and therefore also the best standard for assessing scientific inductions (without necessarily assuming that Bayesianism supplies the *only* standard or means for such inference).

The aim of this paper is to critique, from a Bayesian perspective, the long-dominant view of artificial intelligence as an exclusively symbolic enterprise. My aim being primarily critical, I will not be defending Bayesianism explicitly, except with respect to some of the direct criticisms which have been raised previously within the AI literature. There already is an extensive philosophical literature dedicated to attacking and defending Bayesianism.<sup>4</sup>

I proceed by first reviewing logicism within AI, the view that artificial intelligence is to be achieved exclusively or primarily by the use of deductive inference. Nonmonotonic formalisms have evolved out of logicism’s commitment to symbolic processing under pressure from examples

of ordinary default inferences that lead to retractable (defeasible) conclusions. These formalisms have difficulties with modus tollens type inferences and with specificity, which are reviewed; it is suggested that such difficulties are resolvable via probabilistic inference. In any event, the most promising interpretation of defeasible inference thus far appears to be a probabilistic interpretation. I proceed to consider the lottery paradox, which has long been held to be an insurmountable barrier to probabilistic inference. I demonstrate that it is a *general* problem for inductive inference methods—including nonmonotonic formalisms—and conclude by indicating means for dealing with it within a probabilistic framework.

### 3 The Logician Program in AI

The logicist research program has occupied a pre-eminent position within artificial intelligence—although neither it nor any other paradigm for AI has dominated research to the point of excluding other conceptions. ‘Logicism’ was originally used to characterize the philosophical enterprise of securing our mathematical knowledge in the indubitable bedrock of our *a priori* logical intuitions, as exemplified in the work of Russell and Whitehead in their *Principia Mathematica*. That original logicist program failed in part due to Gödel’s incompleteness result (1931). Logicism in AI, on the other hand, does not propose to reduce mathematics to logic. It proposes, rather, that prerequisite to developing an artificially intelligent system we must find means for representing and reasoning with a vast store of propositional knowledge and that the only viable candidates for representing propositional knowledge are formal logics and their associated formal languages. This view is a not entirely unnatural extension of Church’s and Turing’s research leading to the Church-Turing Thesis. That thesis states that any precise definition of ‘algorithm’ or ‘definite procedure’ will turn out to be equivalent to Turing computability, as did Church’s definition of effective calculability and those of various other people. Assuming the thesis is true, then since Turing computability is defined in terms of universal Turing machines—which are abstract symbol-processing machines—it follows that any algorithm which is sufficient for the production of artificial intelligence (if there be such) is equivalent to some symbol-processing machine. In some important sense, then, symbol processing exhausts the possibilities of algorithmic AI. Hence, the broad support within AI for Newell and Simon’s Physical Symbol System Hypothesis (Newell and Simon 1976). Symbol processing is not, however, exhausted by deduction—for example, heuristic programming is not mere deduction. Nevertheless, logicism draws succour from such reasoning inasmuch as logical languages have considerable representational capabilities (cf. Hayes 1977) and logical inference offers considerable computational power. Logical inference has long appeared to offer the best prospect for producing artificial intelligence by symbolic means.

I can find no fault with the argument above to the effect that algorithmic AI is symbolic AI. What I object to is the idea that logicist, symbolic AI exhausts AI. That is, what I find objectionable are the two exclusionary notions that *only* formal logic has a role to play in symbolic representation and that *only* non-quantitative symbol processing has a role to play in inference. These notions have been coupled with a tendency to consider deduction to be the only rational variety of inference. Notice that it is only within the 1980s that defeasible inference (that is, any inference that is not strictly deductively valid) has been generally acknowledged within the AI community to be legitimate. I shall argue that an exclusive reliance on qualitative inference—including logicism’s rebirth within default logic—is inimical to machine learning and, therefore, to the wider goals of artificial intelligence.<sup>5</sup>

The goal of AI has always been to produce an intellectually autonomous, intelligent, rational agent using a computer as the primary vehicle for intellection. Its inspiration lies in the surprising

computational power (surprising when AI arose, namely during the 1950s) and algorithmic flexibility of the electronic digital approximations of Alan Turing’s universal machine. While computers and computer science have progressed by leaps and bounds since their introduction, their use in achieving the aims of AI has been far less impressive. Hence, there is much talk about “the AI problem.” David Israel (1985) offers two formulations of the AI problem: (I) How do we impart a normal body of commonsense knowledge to a robot? Or, alternatively, (II) how do we get a robot to learn such a body of knowledge? Israel thinks that (I) and (II) are alternative formulations of the same problem. They are not; and the confusion of the two is fundamental.

In his “Programs with Common Sense” (1968) John McCarthy proposed that constructing a program he called the Advice Taker be taken as a central problem for artificial intelligence. The Advice Taker ‘learns’ just by absorbing what the programmer tells it. Theorem-proving is its key to understanding: “We shall . . . say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows” (1968, p. 403). In Israel’s account, McCarthy’s project is to create a “sentential automaton . . . whose abstract data structures are sentences of a formal language and whose interpreter is—or includes—a sound theorem prover for that language. . . . We impart commonsense knowledge to a robot by first formalizing and axiomatizing this knowledge, as completely as we can, and then telling it all to the robot” (1985, p. 429). Or, in another suggestive formulation: “I am urging that we may want to think of *the first-order predicate calculus as a universal machine-language for Knowledge Representation*” (p. 445; his emphasis). Clearly, this program—like philosophical logicism before it—is one that has been deeply impressed by the successes of twentieth-century mathematical logic and axiomatics: it places a theorem prover (whether supplemented or not) at the core of the artificial epistemic agent.

This model of artificial intelligence as the natural outcome of a logic machine has led many to conclude that machine learning is irrelevant to the mission of AI. Those who look to expert systems as a model for AI programs might think that human experts can do our learning for us. Others believe that a key ingredient missing in our AI programs thus far—common sense—can be supplied by axiomatizing some large fraction of what humans know. Neither of these views can be sustained.

### 3.1 Machine Learning as Artificial Intelligence

The first claim is that inductive machine learning is unimportant, for the reason that the general rules that might be so induced can always be extracted from human experts. What AI programming is about then is just deductively applying those rules to initial data to solve some problem or classify some situation. In effect, the AI systems envisaged are just super-calculators doing computations humans already understand, only faster and perhaps more reliably than humans could.

This idea that AI systems just are super-tools is one that Herbert Simon has pushed in his “Why Should Machines Learn?” (1983). He points out there that the learning we observe in humans has various features that are clearly not what we want to emulate in artificial intelligence: the human learning of substantial intellectual tasks is hard, takes many years, is preceded by many other years of dependency upon adults, and when acquired cannot be simply copied and passed around by floppy disk from one human to the next—each human must repeat the hard study on her or his own. Mimicking *these* features is surely not what we want of artificial intelligence, even though they may well be desirable features of programs intended to model human cognitive function. Regarding the latter, Simon is quite prepared to allow that learning

programs are important for computational psychology; but this has no implications for normative AI.<sup>6</sup>

These points are intended to be suggestive, and that is all that they can be. Slowness of acquiring knowledge cannot be essential to learning, and indeed it has nothing to do with the definition of learning Simon himself proposes: “Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively next time” (p. 28). And our current inability to copy the contents of our brains is presumably due to our inability to understand our brains, rather than to anything more fundamental.<sup>7</sup> Simon requires a positive argument for the irrelevance of machine learning.

The positive argument comes in two parts. First, Simon divides machine learning into hypothesis discovery (or invention) and learning; second, he argues that, whereas investigating the automation of discovery is legitimate and important, the learning aspect is irrelevant. Simon’s distinction between hypothesis discovery and learning is similar to the distinction made within the philosophy of science between the context of the discovery or pursuit of hypotheses and the context of their justification. There are some perfectly legitimate grounds for drawing such a distinction. In particular, abductive methods, which are heuristic means of selecting some small subset of explanatory hypotheses out of the (normally) infinite set of possible explanatory hypotheses, are directed at *proposing* hypotheses for pursuit, rather than at accepting or rejecting hypotheses on the basis of evidence. And the ignoring of that distinction has led to some dubious methodologies proposed for science, such as “inference to the best explanation,” which tends to take the abductive step as *both* the first and last step in an induction (cf. Harman 1965). But the legitimacy of this distinction does not imply that Simon is right to exclude discovery from the domain of machine learning. On his own definition he is wrong: the discovery of new hypotheses, unless immediately forgotten or not followed up, will make a difference in how the system subsequently deals with problems similar to the one which led to the discovery—ergo, machine learning. Furthermore, there is little doubt that the contexts of discovery and justification, while distinct, are interrelated—some abductive methods are more fruitful, turn up hypotheses more likely to be true; this is properly reflected in the probability of the hypothesis prior to testing, and Bayes’ theorem tells us that the probability of the hypothesis in the light of the evidence must take that prior probability into account. In short, if machine discovery is legitimate and important, then machine learning is important.

Simon proceeds to downplay the role of interesting learning (relying upon discovery) in favor of the primacy of rote learning, which in an AI program would be nothing more than data entry into a knowledge base. The notions Simon expresses here appear to underwrite much of the goodwill that logicism has received. For example (p. 29): “Most of what we know somebody told us about or we found in a textbook. . . . Most of the things we know were discovered by other people before we knew them, and only a few were even reinvented by us.” Now this claim is not based upon any survey of the jobs to which our different neurons are dedicated. I suppose it is an intuitive response to the difficulties Simon himself experienced pouring over textbooks or trying to teach others. I flatly assert that Simon’s intuitions here are wrong. On the contrary, I claim that the majority of information content that could be found in our heads—had we the tools to find them—is dedicated to keeping track of such mundane things as the way to work, the way back to home, the semantic connections between words, the taste of Coke Classic versus New Coke (or Pepsi), etc. Certainly, if you keep a logbook of how much time you spend studying texts, listening to lectures, or talking to people, then even if you are a professional student, it is most unlikely that the majority of your waking hours will be given over to storing what other people have told you to believe. Whatever the relative role of rote learning in building up our

individual corpora of knowledge, rote learning would have no role whatever were that the only variety of learning available to us. Machine learning and machine discovery are irrelevant to AI only if knowledge is irrelevant.

Well, maybe not. What if all that we demanded of our AI systems was the speedy and reliable computing of solutions to problems that we already understood? This is exactly how Simon construes AI (vs. computational psychology), as having the exclusive goal of getting computers to do difficult computational tasks *so that humans need no longer do them* (p. 27). So Simon implicitly denies the goal of producing an autonomous agent that can learn independently of human support and comfort and that specifically can adapt to an environment that changes in ways unanticipated by its creators. Yet the brittleness of extant AI systems has widely, and rightly, been regarded as a sign of their inadequacy: when confronted with environments at variance with that for which they have been trained or programmed, the performance level of AI systems drops off the table. Humans who perform well within highly restricted domains only are called *idiot savants* for a reason, they lack *general, adaptive* intelligence. The goal of providing a learning, independently adaptive intelligence has virtue that undeniably outstrips the amusement value of observing toy robots rolling about on their own, so long as we do not believe that we have mastered all the problems that our environment has presented, or will present, to us. Indeed, it turns out that the development of systems that can learn about their world to roughly the extent and depth that humans can is a matter of no small practical importance: the inability of expert systems to learn other than by being spoon-fed production rules by “knowledge engineers” is one of the most significant factors impeding their increased use within industry (cf. the discussion of the “knowledge bottleneck” in Hayes-Roth et al. 1983). More generally, the ability to learn is deeply connected with the concept of intelligence. Degrees of learning ability are primarily what we take to distinguish degrees of intelligence among humans, and primarily what we take to distinguish degrees of intelligence across species. If we are to speak properly of artificial *intelligence* at all, we must be prepared to tackle the problem of machine learning.

### 3.2 Naive Physics as Naive AI

The second claim arising from the logicist model for artificial intelligence is that common sense can be attained by axiomatizing much of our own common sense knowledge. This notion has been promoted particularly by Pat Hayes in his Naive Physics Manifestos (1979 and 1985): It is the self-proclaimed task of naive physics to satisfy Israel’s formulation (I) of the AI problem by axiomatizing the various domains of commonsense knowledge. But this idea is based upon precisely the claim that Simon’s argument stumbled over: that rote learning is as much learning as intelligence requires. Although achieving the axiomatizations envisioned for naive physics would be of considerable interest in its own right, that achievement would not obviously be any great advance on understanding either intelligence or common sense. And, I submit, to think otherwise is to be naive. For an example of naiveté, some AI researchers have supposed that what is most striking about the intelligence of a young child is the scope of its factual knowledge about the world, the great range of things it must understand, say, before buying a McDonald’s kids’ meal. Consider the case of IQ tests. These tests frequently fall back on asking people about linguistic exotica. This may be appropriate, but not because intelligence *consists* of the corpus of such knowledge (linguistic or otherwise), but just because that corpus bears symptoms of the processes whereby it was acquired. Certainly, the knowledge prerequisite to buying a kids’ meal is more impressive than the knowledge contained in any AI program. But what is more impressive about *children* is how quickly they can pick up a new language.

In general, the idea that axiomatizing common sense is tantamount to *producing* common sense—no matter how extensive the theories axiomatized—is a totally hopeless concept. It confuses the product of intelligence, our common sense theories of the world, with the processes of intelligence. No matter the scope of such axiomatizations, the robot embodying only that will be unable to generate any new common sense theory for any new environment to which it is introduced. As I have already remarked, an inability to learn even the simplest facts about one’s environment is the antithesis of intelligence and common sense.<sup>8</sup>

Israel tries to defend the importance of the naive physics research program by asking what the criterion of adequacy for our AI robot is, and then answering that it must just be that the robot’s commonsense theory about the world be “informally complete”—which turns out to mean for Israel that the robot successfully incorporates an axiomatized theory adequate for getting about in the world (1985, p. 438). But this just begs the question: this supposed criterion of adequacy is nothing less than the implementation of the logicist program. The *primary* criterion is surely intelligent performance in some broad, not fully predictable, domain of behavior. There may be additional, secondary criteria which refer to aspects of the internal representation of knowledge and inference; but insisting that belief in logicism is a precondition to real AI research is experimentally and theoretically inadequate.

In summary, the production of an autonomous artificial intelligence presupposes an answer to the problem of how to produce an artificial system that can learn about its world. In the sequel I shall examine a number of suggestions about how this might be done without using quantitative reasoning methods; I shall find them lacking. What I am *not* arguing, however, is that these or any other qualitative reasoning methods are without merit. They may well be useful in the sorts of super-tools that I dismissed above: these were dismissed after all not because of their unimportance, for expert systems and other super-tools are very important; they were dismissed simply because they do not offer a model of *autonomous* intelligence, and they should not be misconstrued as doing so. All of my subsequent criticisms are intended to be of like kind: in particular I shall attempt to show that default logic cannot reasonably be construed as offering a model for a learning, inductive intelligence.

### 3.3 The Challenge of Nonmonotonicity

Marvin Minsky pressed some quite different complaints about logicism in an unpublished appendix “Criticism of the Logistic Approach” (1974). One of Minsky’s central objections is that logic demands consistency, since inconsistency plus deductive closure leads to a radical inability to demarcate rational from irrational beliefs; but, since the sets of beliefs of some rational agents (namely, us) are demonstrably inconsistent (as Plato’s Socrates was fond of pointing out), we cannot then use logic to model rational inference. We must employ some inference technique, a control method, which is not so sensitive to inconsistency as are the rules of logic. Minsky goes so far as to claim that consistency is not desirable (1974, p. 76):

I do not believe that consistency is necessary or even desirable in a developing intelligent system. No one is ever completely consistent. What is important is how one handles paradox or conflict, how one learns from mistakes, how one turns aside from suspected inconsistencies.

David Israel (1985) pounces upon this. If consistency is not even desirable, why bother to remove or avoid inconsistencies? But a more sympathetic understanding of Minsky’s words is in order here; we shouldn’t expect Minsky to be so silly as to assert that inconsistency (of all things!) both is and is not desirable. And a perfectly sensible interpretation is available: clearly, what



Minsky means here is that inconsistency is undesirable locally—when a specific inconsistency is found it should be expunged—but the generation of inconsistencies may be an unavoidable side-effect of inference and learning (in the “*developing system*”); that is, an appropriately rich and flexible epistemic agent needs to be capable of reaching *unrecognized* inconsistent states, much as we do. In what follows, I will characterize such an agent as ‘systemically inconsistent’.<sup>9</sup>

The undesirability of local inconsistency, and the long-term unavoidability of systemic inconsistency, are hardly irreconcilable. *We* reconcile them. We manage somehow to learn inductively by coming to conclusions that explain prior facts that puzzle us, while avoiding—or retracting—conclusions that we *know* are likely to be in error. To be sure, there is not universal agreement about this ability. Karl Popper, for one, thought that the tension implicit in our having the twin goals of maximizing information content and minimizing error were irreconcilable and in particular that it posed insuperable difficulties for *any* inductive philosophy. If, for example, we adopt the by now standard way of measuring the information content of  $h$  as the  $-\log P(h)$ , then the probability of  $h$  rises as its content diminishes, and *vice versa*. From this Popper concluded that scientists “have to choose between high probability and high information content, since *for logical reasons they cannot have both*” (1959, p. 120; his emphasis). Popper notwithstanding, the conclusion does not follow from the above inverse relation, as there is no incompatibility between high *prior* content for  $h$  and high *posterior* probability for  $h$ . (The inverse relation obtains for the probability of  $h$  and its content only if they are calculated using the *same* probability function.) When we aim for high content, we aim for theories that go well beyond our initial state of knowledge. That inductive leap makes error unavoidable in the long run. But recognition of that unavoidability is not tantamount to a passive acceptance of inconsistent states nor to a lack of a preference for truth over falsehood. We *assess* our inductive leaps in the harsh light of posterior evidence.

Israel has a more relevant response to Minsky’s inconsistency argument against logicism, drawing a distinction between the use of logic as a representation language and the use of logic *per se*. Israel’s idea is not, of course, that knowledge should be encoded in formal language while formal inference rules be abandoned: he allows full use of formal inference rules. The idea is rather that the logical inference rules are to be embedded in a *broader* control structure. As he says, although a set of inconsistent premises can be used to infer anything at all, that fact does not show that they *should* be so used.

The remaining question is, What is the nature of this broader control mechanism? What I will continue to call logicism below—exemplified by both Israel and McCarthy—displays a strong preference for minimizing non-logical rules of inference and the exclusion of quantitative modes of reasoning.

### 3.4 The Response of Default Logic

Many of Minsky’s objections to logicism revolved around the existence of exceptions, on the importance of *defeasible inference* in human cognition. The point is that we often, if not typically, draw conclusions from what we know or believe that are far more tentative than the premises we start from. Thus, knowing that Tweety is a bird and that birds generally fly, we happily conclude (perhaps) that Tweety can fly. While such inferences may be rational, they are hardly exceptionless.<sup>10</sup>

Allowing exceptions to our accepted rules of inference violates the standard canons of deductive inference: it is placing the center of gravity for our inferential world in some other place than its traditional home, deductive validity.

Unsurprisingly, there has been some resistance to this shift. As David Israel points out

(1980), traditional logics may be employed in nonmonotonic ways—that is, logic alone does not dictate that our knowledge bases must only grow and never contract. Thus, when a new belief is added and an inconsistency arises, we may repair the damage by rejecting one or more premises that gave rise to the inconsistency, including perhaps that new belief. This does not require unsound rules of inference (although, of course, selecting something to reject requires *more* control structure than a theorem prover provides). But what Israel fails to notice is that the Tweety example (and indefinitely many others) *does* employ such unsound inference rules. If we discover that Tweety has a broken wing, then none of the premises we invoked, nor the new belief that Tweety does not fly, is a promising candidate for rejection in order to clean up our belief set. What happened? We *did not* use *modus ponens* in the first place. We used an entirely different kind of inference rule, a default inference rule, whose conclusion can be retracted without impugning the epistemic standing of any of its premises. That is what is strikingly nonmonotonic about default inference: not that premises may be retracted, but that conclusions may be retracted without affecting the premises. (Of course, the latter characteristic does not preclude retracting premises as well, should circumstances warrant it.)

Many (or most) logicians have given up on defending monotonic logic as the primary vehicle for automating intelligence (for a notable exception see Poole 1989). Like McCarthy, they have moved on to formalizing nonmonotonic inference. But it seems appropriate to continue calling them logicians since the approach taken has almost always been to *extend* formal logic, by including new inference rules and axiom schemata. The nonmonotonicity that is characteristic of defeasible inference obliges us to recognize it as a form of inductive inference: if the retraction of a conclusion does not impugn any of the premises used in arriving at it, then the conclusion *ipso facto* is an ampliative one. As I noted earlier, the default inferences addressed by default logic lead from particular to particular, rather than engaging in the testing or validation of general hypotheses. But they are no less inductive for that, and it will be our special concern here whether the logicist refusal to employ the usual quantitative tools for assessing inductions—especially probability theory—leaves them in a tenable position for modeling defeasible inference.

What has come to be called default logic (following Reiter 1980), and which is the most popular AI formalism for effecting nonmonotonic inference, adds default rules of the form:

$$\frac{A : B}{C}$$

Such a rule supports the defeasible inference of C given knowledge of A, so long as B is consistent with what we know. More precisely, if A (conventionally called the *prerequisite*) is a member of K (where K is the corpus of knowledge/belief) and if B (the *justification*) is consistent with K, then defeasibly add C (the *consequent*) to K. If the above rule is named ‘d’, we can use ‘PRE(d)’, ‘JUS(d)’, and ‘CON(d)’ to refer respectively to its prerequisite, justification, and consequent; i.e., PRE(d) = A, JUS(d) = B, and CON(d) = C.

Successive use of default rules, employing intermediate default conclusions as the prerequisites of further rules, can characterize the default import of a state of belief K. We can introduce a formal concept of a default proof based upon this intuitive idea of using a sequence of default rules (cf. Reiter 1980, p. 99).

A default rule for the Tweety inference would be:

$$\frac{\text{Bird (Tweety)} : \text{Flies (Tweety)}}{\text{Flies (Tweety)}}$$

This kind of rule is called a *normal* default because the consequent and justification are identical (i.e., JUS(d) = CON(d)); such rules can be abbreviated in the fashion Bird → Flies. In our

initial epistemic state we do not know of any reason to believe that Tweety cannot fly, so the default rule allows us to conclude that Tweety does fly (if we had known that Tweety cannot fly, the justification  $\text{Flies}(\text{Tweety})$  would not have been consistent with  $K$ ).<sup>11</sup> Assuming we can develop sensible criteria for what is and what is not a good (justified) default rule, then this extension to logic promises to allow us to explore the (default) implications of our beliefs. Cashing in on this promise turns out to have been extremely difficult and has absorbed most of the energy of default theorists. This research program has so far failed to meet the basic criterion of adequacy: to provide a non-ad hoc formalization that (mostly) endorses default conclusions that humans readily accept and does not endorse those that humans strongly reject. This failure has been freely admitted by at least some proponents of nonmonotonic formalisms. (E.g., Reiter 1987, p. 183: “...we know almost nothing about reasonable ways to compute nonmonotonic inferences.” See McDermott 1987 and Etherington, et al. 1991 for interesting discussions of this point.) Defenders will point to the many technical obstacles that stand in the way of producing an adequate nonmonotonic formalism. It may be, however, that many of the technical difficulties have underlying conceptual causes.

## 4 Problems with Default Logic

The general model for an independent intelligence suggested by default logic is that of an agent using default rules to predict events in its world and using observations to confirm or correct these predictions. The ability to perform defeasible inference must be supplemented by complicated perceptual mechanisms and the means to do belief revision, in particular throwing out default conclusions that get in the way of observational facts. This is the logicist view of default logic. Whether or not default logic may be useful otherwise for artificial intelligence, I shall argue that this model of intelligence must be wrong. For this logicist model is of an agent that can only learn about the world what is directly given it by its designers (in the form of magically available default rules) and by perception. There is no ability in prospect for the agent to be able to generalize about its world, to generate and evaluate default rules on its own, without relying upon magic. What are lacking are any useful criteria of adequacy for the default rules themselves. The difficulties posed by the lack of such criteria are clear in the technical development of default logic.

Much of that development has occurred through considering cases of defeasible inference which are intuitively clearly good or bad, and then adjusting the rules of the logic so as to match our intuitive classifications. To take a simple example, consider Figure 1, the “Nixon diamond” (referring to Richard M. Nixon). This is an inference net, where each ‘ $\Rightarrow$ ’ arc represents a universal generalization, ‘ $\rightarrow$ ’ represents a default rule, and ‘ $\neg\Rightarrow$ ’ (or ‘ $\neg\rightarrow$ ’) represents the denial of a universal generalization (or, of a default rule)<sup>12</sup>—in this case, we have the default rule  $\text{Republican}(x) \rightarrow \neg \text{Pacifist}(x)$ . Here we have two potential chains of inference, one leading to  $\text{Pacifist}(a)$  and one to  $\neg \text{Pacifist}(a)$ , where ‘ $a$ ’ designates Nixon.<sup>13</sup> This means that we can use default logic to get two competing, incompatible extensions<sup>14</sup> of  $K$ :  $K' = \{\text{Nixon}(a), \text{Quaker}(a), \text{Republican}(a), \text{Pacifist}(a)\}^*$  and  $K'' = \{\text{Nixon}(a), \text{Quaker}(a), \text{Republican}(a), \neg \text{Pacifist}(a)\}^*$ . But the existence of multiple extensions here is not really a problem. It is true that *we* have grounds for preferring one conclusion over the other, but those are *additional* grounds. Were we confronted with only the information that Nixon is a Republican and a Quaker, then the response that the information is ambiguous with respect to pacifism seems quite legitimate. In other words, it is a plausible interpretation of default logic that only sentences in the intersection of extensions to the knowledge base are to be considered valid default conclusions.

Insert Figure 1 about here

Consider then Figure 2, which is superficially similar to the Nixon diamond. Here again we have two potential chains of inference. Given Whale(a), we can conclude by deduction that Marine-Animal(a) and that Mammal(a). We can then use the default rule to infer Gilled(a) or use *modus ponens* to get  $\neg$  Gilled(a). Intuitively, it is clear that we want to prefer the latter inference over the former. In this case it is no trouble to “adjust” the default logic to correspond to our intuitions: application of the default rule in the first inference would require that  $\{\text{Whale(a), Marine-Animal(a), Mammal(a)}\} \cup K \cup \{\text{Gilled(a)}\}$  be consistent, which is false. The only difficulty here is to ensure that any implementation of the default logic includes heuristic consistency checks that go at least so far as to discover this level of inconsistency.<sup>15</sup>

Insert Figure 2 about here

A kind of statistical syllogism, by contrast, has been the source of serious difficulties for default logic. Consider the following argument:

Most birds are fliers  
Most fliers are insects  
Most birds are insects

The conclusion that Insect(Tweety) can be defeated trivially by pointing out that we know independently that *no* birds are insects. But I can embellish the story to the point where this defeater is unavailable. Suppose that in the biology department some bioengineers have managed to mix together the genes of bees and robins, producing some hideous monster that can fly. It is not inconceivable that we would accept a linguistic convention to describe these things as *both* birds *and* insects. Since there are no great number of these (nor are they typical of birds, etc.), we can represent the inferential setting with Figure 3.<sup>16</sup> Here once more we have competing inferential chains. However, a preference for deduction will not resolve the impasse. Nor is resignation to ambiguity acceptable: the lack of conclusion (incompatible multiple extensions) that one obtains by a direct translation of the arcs as default rules is inconsistent with our clear preference for denying that an arbitrary bird would turn out to be an insect.

Insert Figure 3 about here

There have been a number of responses to this kind of problem (known as the ‘specificity’ problem). One way of forcing a preference is to rewrite the default rules to name explicitly those exceptions which defeat them: instead of the “normal” rule (a) use the “semi-normal” (b):

(a)  $\frac{\text{Flier}(x) : \text{Insect}(x)}{\text{Insect}(x)}$     (b)  $\frac{\text{Flier}(x) : \text{Insect}(x) \wedge \neg \text{Bird}(x)}{\text{Insect}(x)}$

Touretzky (1984, p. 107) notes that moving to semi-normal rules forces the complexity of default rules to increase as the knowledge base grows. But things are even worse than that. Many of the concepts of ordinary language are not resolvable into a clean set of necessary and sufficient conditions, as Wittgenstein pointed out for the concept of *game* (1958, paragraphs 68-70). Any proposed set of sufficient conditions will have exceptions. For example, whatever criteria we propose for an activity being a game, we can always invent a new game using new kinds of activity (unless the sufficient conditions offered circularly make reference to something being a game, when we have no interesting analysis of the concept). But if that is so, there is no reason

to believe that the program of implementing autonomous intelligence using semi-normal rules can succeed: as our concepts have this open texture there is no point at which we can claim to have a complete list of exceptions. The default “logician” shall continue to need magical assistance to provide truly intelligent editing of its semi-normal rules.

Another attempt to deal with the birds and the bees has been to point out that the path from Bird(Tweety) to  $\neg$  Insect(Tweety) is shorter than that leading to Insect(Tweety) (e.g., Fahlman’s NETL works this way; see Fahlman 1979). Unfortunately, this shortest path criterion is overly sensitive to our language and state of knowledge. For example, if we added a node between Bird and  $\neg$  Insect, such as Vertebrate, then the shortest path criterion fails to yield an answer, even though we know that being a vertebrate does not predispose anything to being an insect.

Touretzky (1984) presents an alternative approach to enhancing default logic so as to avoid indecision, or the wrong decision, in this case. He introduces a partial ordering on normal defaults by which he can prefer arguments based upon preferred default rules. A rule  $d_i$  is preferable to  $d_j$ , written (perhaps counterintuitively) as ‘ $d_i \ll d_j$ ’, if and only if

- (a) there is a  $d_k$  such that  $\text{PRE}(d_k) = \text{PRE}(d_i)$  and  $\text{CON}(d_k) = \text{PRE}(d_j)$ . For example, if we have a chain of defaults  $A \rightarrow B \rightarrow C$ , then  $A \rightarrow B \ll B \rightarrow C$ , because  $\text{PRE}(A \rightarrow B) = \text{PRE}(A \rightarrow B)$  and  $\text{CON}(A \rightarrow B) = \text{PRE}(B \rightarrow C)$ . I.e., the initial link of a chain is preferred.

or

- (b) there is a  $d_k$  such that  $d_i \ll d_k$  and  $d_k \ll d_j$ . For example, if we add one link to our chain in (a) we get  $A \rightarrow B \rightarrow C \rightarrow D$ ; by clause (a) we get both  $A \rightarrow B \ll B \rightarrow C$  and  $B \rightarrow C \ll C \rightarrow D$ , so by clause (b)  $A \rightarrow B \ll C \rightarrow D$ . (I.e., clause (b) allows us to prefer the earlier links in a chain of defaults to subsequent links.)

In another case of interest, if we are given the three rules:  $A \rightarrow B$ ,  $A \rightarrow C$ ,  $B \rightarrow D$ , we get the preference  $A \rightarrow C \ll B \rightarrow D$ , because  $\text{PRE}(A \rightarrow B) = \text{PRE}(A \rightarrow C)$  and  $\text{CON}(A \rightarrow B) = \text{PRE}(B \rightarrow D)$ . Viewing the graph below, we can see that a one-link chain with a common head gets preferred.

Insert Figure 4 about here

Now in the default representation of Figure 3 we have:

- $d_1: B \rightarrow F$
- $d_2: F \rightarrow I$
- $d_3: B \rightarrow \neg I$

As I mentioned, this representation leads to multiple extensions in unvarnished default logic. But we can use the ordering of the default rules to impose an ordering on default proofs as well. In this case, we have the rule preferences  $d_1 \ll d_2$  (by clause (a), letting  $d_1$  itself be  $d_k$ ) and  $d_3 \ll d_2$  (by clause (a), letting  $d_1$  be  $d_k$ ). The default proof yielding Insect(Tweety) uses the following sequence of defaults:  $p = \langle d_1, d_2 \rangle$ ; the alternative proof of  $\neg$  Insect(Tweety) uses  $p' = \langle d_3 \rangle$ . Touretzky’s rule is to order default proofs according to the order of the last default rules employed; since  $d_3 \ll d_2$ , it follows that  $p' \ll p$ . Hence, only the conclusion  $\neg$  Insect(Tweety) is permissible.

This seems to be a real improvement over normal default logic. As Pearl points out (1988, p. 472), Touretzky’s rule forces our knowledge about subsets to take precedence over our knowledge

about supersets, which is surely what we want. Returning to the case of Tweety the Penguin’s flying abilities, if we add a default rule to represent the subset relation Penguin  $\subset$  Bird ( $d_2$  below), we get Figure 5 or:

$$\begin{aligned} d_1: B &\rightarrow F \\ d_2: P &\rightarrow B \\ d_3: P &\rightarrow \neg F \end{aligned}$$

Insert Figure 5 about here

If we take  $d_2$ , the rule expressing the subset relationship, as  $d_k$  in clause (a) we find that the default specific to that subset (Penguins) is preferred to the default applying to the superset (Birds); i.e.,  $d_3 \ll d_1$ .

Unfortunately, Touretzky’s system is not a general answer to the problem of selecting between competing chains of inference. It succumbs to the same linguistic dependency that afflicted the shortest path criterion.<sup>17</sup> Taking the same statistical syllogism above, but adding the node Vertebrate between Bird and  $\neg$  Insect, we get Figure 6, or the following four default rules:

$$\begin{aligned} d_1: B &\rightarrow F \\ d_2: B &\rightarrow V \\ d_3: F &\rightarrow I \\ d_4: V &\rightarrow \neg I \end{aligned}$$

Insert Figure 6 about here

Inspection reveals that (aside from the reflexive cases, which can be induced by adding the dummy defaults of the form  $X \rightarrow X$ ) only the following four preference relations can be generated by clause (a):  $d_1 \ll d_3$ ,  $d_1 \ll d_4$ ,  $d_2 \ll d_3$ , and  $d_2 \ll d_4$ . Clearly, we cannot get either  $d_3 \ll d_4$  or  $d_4 \ll d_3$  through clause (a); the former would require the default Flier  $\rightarrow$  Vertebrate and the latter Vertebrate  $\rightarrow$  Flier, neither of which makes sense. Clause (b) just allows us to connect directly the ends of chains in a preference order. In short, we cannot relate by preference  $d_3$  and  $d_4$ , so Touretzky’s system gets stuck in the same way as Reiter’s default logic, with no way to prefer the rule that birds are not insects over the rule that birds are insects.

In general, the idea of assessing the strength of non-deductive chains of reasoning using only syntactic features of those chains appears to be of dubious merit. While it is clear from Touretzky’s work that aspects of those chains are relevant—in particular, subset relations that hold across chains are surely relevant—it is not believable in general that syntactic features of those chains will serve to fully determine the inferential relevance of the beginning to end nodes. From a Bayesian perspective, default logicians are attempting here to measure and compare the conditional probabilities  $P(\text{end-node} \mid \text{beginning-node})$  on the basis of strictly *qualitative* features of an inference net. But having stripped quantitative representations from the network, it is not reasonable to suppose that the remaining features alone could be sufficient to adequately approximate Bayesian reasoning. The Vertebrate example occasions the failure of Touretzky’s preference system. And this argument suggests that any qualitative extension to that system must run up against some other counterexample somewhere.

## 4.1 Modus Tollens

Given that default rules allow us to perform a defeasible kind of *modus ponens*, one naturally wonders whether the inference can work in the reverse direction in a kind of *modus tollens*; that is, are we justified in concluding, given only the fact that Tweety does not fly, that Tweety is not a bird? Certainly, the default logic developed by Reiter does not endorse such reasoning. Nor does it appear that such reasoning was intended to be supported.

A defense for a refusal to support *modus tollens* arguments has been presented by Dubois *et al.* (1985).<sup>18</sup> They describe an example intended to show that allowing *modus tollens* would lead to fallacies. Suppose we adopt the rule that if John attends a meeting, then Bob does not; or for short,  $\text{At-Meeting}(\text{John}) \rightarrow \neg \text{At-Meeting}(\text{Bob})$ . Surely, this cannot *by itself* support the contrapositive rule presupposed by *modus tollens*,  $\text{At-Meeting}(\text{Bob}) \rightarrow \neg \text{At-Meeting}(\text{John})$ . We can give an interpretive story that rules out the latter, in fact. Suppose John is Bob's boss and that he arranges and attends all of Bob's meetings. Then not only is the contrapositive false, its universal denial,  $\text{At-Meeting}(\text{Bob}) \Rightarrow \text{At-Meeting}(\text{John})$ , is true. This is, of course, consistent with the vast majority of John's meetings being unattended by Bob, which would be normal for a manager. This example makes its point. A default rule cannot by itself support a *modus tollens* argument, unlike universal generalizations. It is curious, however, that the example used to defend this feature of default logic relies entirely on a *probabilistic* insight (see Figure 7).<sup>19</sup>

Presumably, the conclusion must be that the contrapositive rule must stand or fall on its own merits. But this leaves us with the awkward question, What merits are those?

Insert Figure 7 about here

There surely are cases where a default *modus tollens* should work. There must be cases where, whatever those merits are, they are sufficient for allowing both directions of inference. Indeed, the birds-and-bees is one such case, since despite the occasional biochemical experiment, discovering that something is an insect would strongly predispose us toward concluding that it is not a bird. For a more homely example we can turn to the early 1991 baseball season. Consider the rule, when Roger Clemens pitches, the Boston Red Sox win, or  $\text{RC} \rightarrow \text{Win}$ . We had pretty good evidence (if limited to the early season) that this was a good rule. Now suppose that it was Clemens' turn in the pitching rotation yesterday (sometime during early 1991). Then I can conclude that Boston won yesterday. But it could be that I heard from someone that the Yankees beat Boston 15-2. It would seem perfectly rational to conclude that Clemens did not pitch in his turn. Default logic will not lead us to that conclusion, however. The fact is, if we are given

- (a)  $\text{RC} \rightarrow \text{Win}$
- (b)  $\text{RC}$
- (c)  $\neg \text{Win}$

then, contrary to the logicist view of default logic, *any* of the following conclusions are possible (except for the second case, it is also assumed that  $\text{Win}$  has been retracted):

- (1)  $\neg \text{RC}$  (Clemens missed his turn)
- (2)  $\text{Win}$  (Our informant was unreliable)
- (3)  $\neg [\text{RC} \rightarrow \text{Win}]$  (Clemens' arm is dead)
- (4) (a)-(c) are OK (The default conclusion was in error)

There is a perfectly good story potentially available for each of these conclusions. Which conclusion we accept ought rationally to depend on which story we think is most likely true. A

mechanical retraction of the default conclusion, leaving the other options unconsidered, would display a severe inability to assess either the quality of the evidence or the validity of the rules. If we learn (b) or (c) through a news report, we justifiably will give it more credence than if we were to learn the same via the off-hand remark of someone at a bar. Likewise, we can have direct or indirect evidence to support or weaken the rule (a). *There is nothing in default logic that addresses any of these issues*—yet they are hardly inconsequential: the default retraction of option (4) is senseless if any of (1)-(3) are well supported conclusions. In short, we need means to assess the strength of support for all of the premises and the conclusion of a default inference, as well as the support for the default rule itself. These are just the kinds of questions that Bayesian methods have always attempted to address.

## 5 Some Possible Semantics for Default Rules

This brings to the fore the question of the relation between default rules and the corresponding conditional probabilities. Sometimes the question has been put: Aren't default rules just probability statements in disguise? That probabilistic formulation is easily defeated, if it is conceded that default rules are rules and that rules are neither true nor false. But this “victory” for logicism is both trivially easy and trivial. It only pushes the problem back one step: Are there any such rules? I prefer an alternate way of re-raising the issue: Is not a sufficiently high conditional probability a necessary condition for the acceptability of some (or many) default rules? I cannot here provide a full defence of the positive answer, but I shall outline such a defence, while criticizing two popular alternative proposals for giving default rules a semantics: typicality relations and the adoption of rules by convention.

### 5.1 Typicality, or Most Birds Don't Fly

J. Terry Nutter (1990), objecting to the probabilistic understanding of defeasible inference, points out that there are times when in fact most birds do not fly, namely during the spring when most birds are unfledged. (This can surely be doubted on a variety of grounds, including its hemispherical bias; but we can be charitable here, inasmuch as it *could* be true sometimes and in some linguistic contexts.) Presumably, we should continue to accept and use the default rule  $\text{Bird} \rightarrow \text{Flies}$ ; otherwise, her point would not be an objection to the probabilistic interpretation of defaults, but a capitulation. But how can the statistical fact and the contrary default be made compatible? Nutter's answer is that the rule expresses a typicality claim: since the “typical” bird flies—and Nutter cites unsurprising psychological evidence that people do treat flying birds as more typical than flightless birds—the rule is a good one.<sup>20</sup> But good for what? It is one thing to be good for, say, descriptively representing the way most people think, or for explaining associative priming between concepts, but it is quite a different thing to say that the rule is good *as a normative guide for inference*, which is all that we are concerned with here. Nutter's position applied to *normative* AI would amount to this: so long as typicality endorses it, infer what you know is most likely false.<sup>21</sup>

The object of inference to a conclusion is normally to build a model of the world that is maximally useful in explaining, predicting, and maneuvering within that world. It is (nearly) universally accepted that contradictions and errors are not good *in themselves*, and so we should not endorse policies that are more likely than alternative policies to lead to error, when there are no significant compensating virtues. It seems, therefore, that we must reject Nutter's policy of preferring improbable, “typical” conclusions over probable conclusions.

At first glance, my argument may appear circular. That is, the question before us is just



whether default rules, when legitimate, are backed by probabilities. The counter has been made that the semantics of default rules are to be understood in terms of typicality relations and that, therefore, where there is a conflict between typicality and probability, the latter should go. I reject this argument because in such cases typicality is not backed by the requisite probabilities. Of course, I could play *that* game with any proposed counterexample to probabilism and never be shaken off of my carousel. But circular arguments are not very convincing to unbelievers!<sup>22</sup> However, my argument in a broader context is not circular. In rough outline it goes as follows. There are independent arguments for believing that probability claims support inference. These are specifically the various Dutch book arguments—demonstrating that by violating probability theory you render yourself open to *guaranteed* losses from supposedly *fair* bets—and the good sense that the Bayesian analysis makes of confirmation theory (see Howson and Urbach 1989 on both counts). These arguments, to support inference and not just deductive reasoning about probabilities, need to be supplemented by arguments for probabilistic acceptance, which I do in part below. The sum of these arguments, then, supports the claim that there are good independent reasons for believing that probabilistic inference has wide applicability as a normative guide to commonsense and scientific inference. No such arguments have been forthcoming to support the rationality of inference based on typicality claims *except to the extent that such claims are also supported probabilistically*. There is no Dutch book argument for typicality-based inference. What has been offered most commonly as supporting typicality-based inference is a collection of cases which intuitively are cases of rational inference and which reflect our typicality judgments. But, given that we have independent reason to believe that probabilities *can* support such inference, if the probabilities in these cases are shown to run in the right direction, then there would be no additional inferential work left over for typicality to do—and so, no reason to believe that typicality supports normative inference.

The remaining question is whether in Nutter's case the probabilities *do* flow in the right direction. The answer depends, of course, on whether one believes that—under the particular circumstances—the default rule is right. The psychological evidence cited does not support Nutter's case so far as *normative* inference is concerned. The *intuition* that the rule is fine as is may perhaps withstand the point that persistent inference to known or probable falsehoods will do severe epistemic and economic damage, but it is certainly not an intuition that I share. (I would be happy to make bets with Terry Nutter come springtime about birds flying, assuming that she will put her money where her defaults are!)

Other than outright rejection of the default rule, there are at least two other responses to Nutter's example that, depending upon context, may be reasonable: (a) It could be that conversational implicature—the implicit conventions governing some context of discourse (cf. Grice 1975)—sometimes rules out certain non-standard cases: i.e., if unfledged birds were meant to be within the domain of discourse, they would have to be explicitly identified. If unfledged birds are not even in the domain of discourse, then the conditional probability for flight will be quite high all year around. Thus, both the conditional probability and the default rule point in the same direction. (b) Within a specific context, the rule may refer to *species* rather than to individuals, in which case again the conditional probability will remain high. Clearly this will not always be the right interpretation, for sometimes we talk about species and sometimes we talk about individuals.<sup>23</sup>

## 5.2 Convention, or Birds Shall Fly

Nutter (1990) extends the above objection to the probabilistic account of default inference by reference to the many-splendored nature of defaults: It is not merely the case that probabili-

ties cannot account for typicality-based generalizations, but there is a large variety of default-supporting generalizations; probabilistic generalizations are only one such kind, the (or many, or some) other kinds cannot be accounted for in terms of probability (cf. also Reiter 1987). All of the following varieties have been discussed in this connection at some point:

- Typicality Generalizations: Birds fly.
- Fuzzy Generalizations: Basketball players are tall.
- Causal Generalizations: Drunks have poor reactions.
- Linguistic Conventions: Birds fly.
- Normative Principles: Consider defendants innocent until proven guilty.
- Methodological Rules: Assume things are as they appear to be.

Each of these appear to support defeasible inference. There may be some dispute about the merits of the examples I have selected, but if they are accepted, then in each case the conclusions they endorse are surely nonmonotonically retractable on the basis of conflicting additional evidence. It is not necessary for Bayesian AI to claim that *all* of these varieties of rules must be amenable to probabilistic analysis; nor do I claim that. But I do claim that the two categories most popularly drawn upon by default logicians are so amenable, namely typicality generalizations—dismissed above—and linguistic conventions.<sup>24</sup>

John McCarthy has proposed (1986) that we interpret default rules as expressing conventions that govern discourse. No such convention, for example, “requires that most birds fly. Should it happen that most birds ... cannot fly, the convention may lead to inefficiency but not incorrectness” (p. 91). This stands on its head my use above of Gricean implicature to rule non-fliers out of the domain of discourse. My point there was that in that case implicature can make the statistical facts break the right way. But McCarthy is surely also right that our conventions do not *guarantee* that the statistical facts will go the right way—and that the interesting question is what to say about cases where the two diverge. If in fact most birds do not fly, then maintaining the default rule in the face of contrary statistical evidence will obviously lead to inefficiency: we will have to nonmonotonically repair a good deal of inferential damage. But then, just as obviously, McCarthy is wrong to say that the rule will not lead to incorrectness. He is probably relying here on the trite point that the *rule itself* is not false and ignoring the fact that *most conclusions* reached by following that rule *are false*. One can ignore the falsehoods and stress the inefficiency only on the implausible assumption that all the falsehoods will be uncovered and repaired without serious harm being done. However, on precisely the same grounds that applied to typicality-based rules, the statistical facts require us to revise any convention-based default rules that go astray: ignoring those facts is known to lead not just to inefficiency, but also to wrong explanations and predictions, and in general to a wrong model of reality. A minimal, regulative concept of rational norms constrains our conventions to reflect *accurately* “the statistical reality that compelled the design and use of these conventions,” as Pearl puts it (1988, p. 478).

The argument from many-splendored defaults appears to have a good point to make, that there is rule-based defeasible inference that cannot be explained or modeled probabilistically. Yet the two examples that have been touted most prominently fail to make that case: typicality generalizations and linguistic conventions appear to support default inference only when the corresponding conditional probability backs that inference. Adding in causal generalizations and foundational principles, these are just the kind of generalizations needed to build an empirical

world view, and so provide an appropriate platform for building an autonomous, Bayesian intelligence.

## 6 Lessons from the Lottery Paradox

We have seen that there are serious difficulties in the way of providing a purely qualitative account of inductive learning as defeasible inference. However, there has apparently been a brick wall lying in the path of any alternative attempt to represent inductive learning using probabilistic acceptance: the lottery paradox. Roughly put, the lottery paradox points out that a simple identification of the acceptability of an inductive conclusion with high probability appears to confuse a low chance of winning a fair lottery with *no* chance of winning it, leaving inexplicable how a lottery could ever be won.

For the last thirty years, since Henry Kyburg, Jr. published it in (1961), many philosophers and AI researchers have thought that the lottery paradox establishes the vacuity of probabilistic inference by showing that it has an unwholesome proclivity to degenerate into out-and-out incoherence.<sup>25</sup> But that is a misunderstanding of the paradox, as Kyburg has argued himself. The immediate reaction by various inductive logicians, such as Hintikka and Hilpinen (1966), was to evade the paradox by imposing strong restrictions upon acceptance, beyond any quantitative confirmation level. But Kyburg, in his paper “Conjunctivitis” (1970), demonstrated that these systems suffered a fate quite as bad as succumbing to the paradox: in effect, the only statements that ended up being acceptable were the grand conjunctions of the observable evidence. Being incapable of *any* generalization, indeed of any conclusion going beyond the evidence, these systems suffered a total inductive collapse. Qualitative systems retreat even farther from probabilistic acceptance. However, the paradox causes trouble for *any* system, qualitative or quantitative, that would attempt induction—that detaches an ampliative conclusion from its context (premises). Indeed, the trouble it causes for qualitative systems is the worse trouble—appearing to be irreparable—while the trouble posed for a Bayesian system with acceptance is at least manageable in principle. I will take John Pollock’s OSCAR (*How to Build a Person* 1989) and default logic for case studies. The response of the former to the lottery paradox is again one of inductive paralysis—a paralysis that spreads throughout the system; the response of default logic is little different.

The lottery paradox put more precisely requires a number of assumptions to get started. Perhaps most prominently, it needs a probabilistic rule of acceptance. An acceptance rule is supposed to warrant the belief in the conclusion of an inductive argument when that conclusion achieves some sufficiently high level of probability. Given an acceptance level of 0.9, say, and a million participants in a fair lottery, then I am warranted in concluding that my own ticket *i* will not win; i.e., if we call this conclusion ‘ $\neg\Phi_i$ ’, from the argument that shows  $P(\neg\Phi_i) = 1 - 10^{-6}$  (i.e., 0.999999), I can inductively conclude that  $\neg\Phi_i$ . While this may seem all right, it is clear that we can draw the same conclusion for every player in the lottery. Hence,  $\neg\Phi_1, \dots, \neg\Phi_{1,000,000}$ . Having accepted one million assertions of failure, we can then conclude both

$$\bigwedge_j \neg\Phi_j \quad (\text{by conjunction})$$

and

$$\neg \bigwedge_j \neg\Phi_j \quad (\text{by fairness of the lottery})$$

that is, both that no one will win the lottery and that someone will win the lottery.

This argument has been thought to provide a conclusive reason—in the form of a *reductio ad absurdum*—to reject probabilistic acceptance and to turn instead to qualitative knowledge representations.<sup>26</sup> But the paradox turns on more than acceptance: as Kyburg pointed out (1970), the paradox requires other assumptions, including at least the following.

1. **The Direct Inference Principle.** If the frequency of property  $Q$  in population  $P$  is  $r$  and if an object  $a \in P$  is randomly selected, then the probability that  $a$  is  $Q$  is  $r$ .<sup>27</sup>
2. **The Weak Deduction Principle.** If a statement is accepted, then its direct consequences are acceptable as well.
3. **The Weak Consistency Principle.** No self-contradictory statement is acceptable.
4. **The Conjunction Principle.** If two statements are accepted, then their conjunction is acceptable.

Kyburg took his paradox not to require a rejection of acceptance, but as an argument from the principles of weak deduction and weak consistency to the inadequacy of the conjunction principle. The conjunction principle entails the equivalence of weak deduction and deductive closure. And it also forces the collapse of weak consistency, or any slightly stronger consistency requirement, into the strong consistency principle that all finite conjunctions of our beliefs be self-consistent. To some, the strong consistency principle is more obviously objectionable than probabilistic acceptance: minimally, we have no working examples of strongly consistent intelligences. It is clear enough that the conjunction principle is as central to the lottery paradox as is acceptance.

## 6.1 The Collapse of Collective Defeat

The lesson that John Pollock draws from the lottery paradox is not that probabilities provide no reason to infer, but rather that what we have here is a case of collective defeat—that is, *none* of the conclusions of the form  $\neg\Phi_i$  may be drawn because they collectively defeat each other. It is Pollock’s stated objective to provide an account of inductive, defeasible reasoning and thereby to facilitate the construction of a real artificial intelligence. He believes that this objective can be fulfilled by laying down qualitative principles of reasoning, of which a principle of collective defeat is fundamental. That principle states roughly that if we have a set of defeasible conclusions where for *each* conclusion there is an argument from the other conclusions (and background knowledge) to its negation, then, regardless of how long or complicated such defeating arguments may be, none of the conclusions is warranted.

Specifically about the detached  $\neg\Phi_i$  in the lottery paradox Pollock says (1987, p. 494; my emphasis): “Intuitively, there is no reason to prefer some of the  $[\neg\Phi_i]$  over others, *so we cannot be warranted in believing any of them unless we are warranted in believing all of them*. But we cannot be warranted in believing all of them,” because of the inconsistency. This is intended to be an application of the principle of collective defeat; I shall call it Pollock’s Rule. Pollock’s Rule can be seen as derivative from the conjunction principle—and viewing it in this light perhaps makes it more plausible than otherwise. The contrapositive of the conjunction principle is: if a conjunction is not acceptable, then not all of its conjuncts are acceptable. Of course, to strengthen this contrapositive to allow, without restriction, the exclusion from our corpus of belief of *all* of those conjuncts, as in Pollock’s Rule, would be absurd. But Pollock’s Rule is restricted to cases where those conjuncts themselves have equal support. By the contrapositive conjunction principle we must exclude some of those conjuncts; by equal support we cannot prefer one conjunct over another; therefore, we appear to be constrained to exclude them all.

This suspension of belief in the face of the lottery paradox has been found plausible before (e.g., Cohen 1983, p. 249, Stalnaker 1984, pp. 91-92, Levi 1983, p. 255, and even Perlis 1987, p. 188). But Pollock’s Rule, however plausible it sounds, fails to support a workable concept of justified belief and therefore it fails to provide an adequate response to the lottery paradox. This can be seen by observing that a reasonably close analogy can be drawn between the lottery paradox and everyday reasoning situations. That is, one can “lotterize” just about any inductive inference problem, and so, if using Pollock’s Rule, one will *almost always* be constrained to indecision, even concerning the most ordinary, dull, unobjectionable inferences.

Consider a sequence of  $n$  coin flips of a fair coin. Any such sequence has a probability of occurrence of  $(1/2)^n$ , which is, of course, a very low probability for large  $n$ . It seems, in advance of flipping the coin, we should be able to state that the sequence that in fact turns up is highly improbable. And so, we would be in a position to infer that the actual state of affairs would not occur. But, naturally, we could say the same about any other sequence. Since *some* sequence must occur, by Pollock’s Rule we cannot infer any of these conclusions. All that we need to trigger Pollock’s Rule here, or anywhere, is a partition of the outcome space such that each member of the partition has the same low probability of occurrence.

We might wonder why we should be concerned about drawing inductive conclusions within such artificial environments as lotteries or  $n$ -length sequences of coin flips. The answer is that we must be concerned about it, because we can *typically* partition inductive inference problems in the way needed by Pollock’s Rule. Consider the practical problem of whether or not to take an umbrella (for fear of rain, not sun) out on a walk tomorrow in the Sahara desert. We can partition the possible states of the weather so that each member of the partition has roughly the same, low, probability of occurring. Any continuous measurement scale that applies to the weather will do, for example the temperature. Thus, it may be that the probability of its being sunny and the high temperature being within 1/100th of a degree of 45 degrees Celsius is the same as the probability of its being rainy and the high temperature being within 3 degrees of 30 degrees. It is always possible to generate some partition with the cells having roughly equal, low probabilities of occurrence (see Figure 8). One minus this low probability is greater than our acceptance level, by stipulation; that is, we will create a fine enough partition that the probability of any member of the partition is sufficiently low. We are inclined to say then of any such event that it will not occur, and so conclude, for example,  $\neg R_{30}$ . But again, there *will be* weather tomorrow in the Sahara, so by Pollock’s Rule we are held to indecision.

Insert Figure 8 about here

Clearly, we can play the partition game with just about any inductive inference problem. But perhaps that does not seem so bad; after all, the only conclusions we are obligated to avoid here are the denials of very particular states of affairs. And deciding whether we need umbrellas tomorrow does not hang on the temperature, but on the precipitation. That is, it seems open to us to forge ahead with the ordinary, boring conclusion that it will not rain tomorrow in the Sahara ( $\neg R$ ) and so that we do not need our umbrellas. But there is a problem with this line of thought: if we can conclude that it will not rain tomorrow, then surely we can conclude—deductively this time—that it will not rain tomorrow with the temperature near 30 degrees (that is,  $\neg R \vdash \neg R_{30}$ ; see Figure 9). Since any conjunction deductively implies its separate conjuncts (*this* conjunction principle is not in doubt!), if we are obliged to *refrain* from inferring some conjunct ( $\neg R_{30}$ ), we are equally obliged to refrain from inferring any conjunction containing it ( $\neg R$ ). Therefore, we cannot conclude that it will not rain tomorrow—we cannot conclude  $\neg R$ ; in general, we cannot reject any set of events for fear of rejecting its component subevents.<sup>28</sup>

Likewise, since the affirmation of a set of events implies the rejection of the events excluded (since  $S \vdash \neg R$ ), we also—absurdly—cannot affirm that it will be sunny tomorrow, no matter how likely that may be (short of certainty). *In the end, we cannot draw any inductive conclusions.*<sup>29</sup>

Insert Figure 9 about here

## 6.2 Diagnosis

As I argued above, Pollock’s Rule is best seen as derivative from the conjunction principle—it is a specialized version of the contrapositive conjunction principle. But why should we believe the conjunction principle? It’s clear that if we believe that probabilities provide inductive warrant, then the conjunction principle cannot be right: the probability of a conjunction of independent non-trivial propositions is *guaranteed* to be less than the probability of any conjunct. The conjunction principle requires us to ignore this fact. Regardless of your attitude toward inductive probabilities, the conjunction principle is inimical to *any* variety of induction: any inductive conclusion will be less certain than its premises; those uncertainties tend to *accumulate* through conjunction; the conjunction principle requires us to ignore this fact. (And, of course, these facts remain when the inductive support for each conjunct is equal, which is Pollock’s special case.) It therefore should be unsurprising that inference systems founded on the conjunction principle, including Pollock’s, end up rejecting induction.

In short, Pollock’s attempt to dispatch the lottery paradox using a qualitative system endorsing the conjunction principle suffers from the very same defect that Kyburg identified twenty years ago in quantitative systems endorsing the conjunction principle: it eliminates induction. A philosophy which endorses the conjunction principle, therefore, can hardly serve as a framework for developing an artificial intelligence that learns inductively about its world.

Perhaps this conclusion will surprise Mr. Pollock: in introductory moments, he quite sensibly urges that “a reasonable epistemology must accommodate both” deductive and inductive reasoning (1987, p. 481). One plausible reason that Pollock and others have been led to the extreme of rejecting inductive inference is an exaggerated concern with the *normativeness* of inference rules. Although I think naturalized epistemology goes too far in rejecting normative concerns, it is also possible to go too far by imposing utopian standards. Pollock is explicit in stating that his theory of warrant attempts to capture what an *ideal* reasoner would be justified in believing. His ideal reasoner is simply “unconstrained by time or resource limitations” of any kind (1989, p. 127). But it is not necessarily true that because such a reasoner is somehow an idealization, we or our AI systems can somehow approximate it: it is not clear how, or that, an epistemology for God has anything much to tell us about human or artificial epistemology. In particular, it may be plausible to those who have such intuitions that the (presumably infinite) conjunction of everything justifiably believed by God is itself justified for Him, but this is not merely implausible for humans, of humans it is known to be false.

The preface paradox (a relative of the lottery paradox) is a nice statement of our limitations: even though we may be justified in believing each individual sentence of a book that we write, we would also be well justified to insert into the preface the statement that at least one (other) sentence in the book is false. What we require is an epistemology of fallible beings, an epistemology that can deal with error and inconsistency. The impossibility of avoiding error was, after all, precisely the original stimulus within AI for investigating defeasible inference. To assume simultaneously that errors will never occur—as is required by the conjunction principle—is somehow inconsistent. It is high time that we recognized the global inconsistency that plagues the systems we are and the systems we build, and stop demanding that the next step in building

a person be to scale up from human to godlike proportions.

### 6.3 The Retreat of Default Logics

The lottery paradox is not just an artifact of probabilistic systems, it is symptomatic of potential troubles in any variety of ampliative, non-deductive system. It also gives rise to difficulties in default logic, for example.<sup>30</sup> We have to assume *some* applicable criterion of acceptability for default rules, so let us say a default rule is acceptable if it expresses a typical property of a class of objects and, for concreteness, that this means that when one learns how to classify objects of that class (e.g., by ostensive definition) the property in question is usually to be observed in the given exemplars (if ‘usually’ suggests that I am begging some question, just substitute ‘almost always’ and rewrite the example below using the number of bird species we in fact encounter). Then we can suppose that in a given linguistic community there are three species of birds: Eagles, Penguins, and Hummingbirds. If these three are all equinumerous and roughly evenly distributed, no *species* will be typical of birdhood; furthermore, it may well be that the following three defaults are acceptable:  $\text{Bird} \rightarrow \neg \text{Eagle}$ ,  $\text{Bird} \rightarrow \neg \text{Penguin}$ ,  $\text{Bird} \rightarrow \neg \text{Hummingbird}$ . If we further assume it to be *known* that there are only these three varieties of bird, then inconsistency threatens when one is presented with an arbitrary bird Tweety. The inconsistency threatens, but does not actually arrive. If we use two defaults to conclude, say, that  $\neg \text{Eagle}(\text{Tweety})$  and  $\neg \text{Penguin}(\text{Tweety})$ , the rule for using defaults will prevent us from concluding that  $\neg \text{Hummingbird}(\text{Tweety})$  because that conclusion would be inconsistent with our current (new) belief set. This way of evading inconsistency is not especially meritorious, however. It leaves the default reasoner with some highly peculiar tendencies of thought. In the given case, it will have defeasibly concluded that Tweety is a Hummingbird, since it can derive that from its default conclusions. But that conclusion came on the basis of no information about Tweety *at all*, beyond its being a bird. Should the default reasoner, for example, always apply default rules in the same order, it will behave as though it has the rule  $\text{Bird} \rightarrow \text{Hummingbird}$ —even though it explicitly has the rule  $\text{Bird} \rightarrow \neg \text{Hummingbird}$ ! Indeed, what conclusions it arrives at in situations of uncertainty will depend upon the arbitrary matter of the order in which it considers its default rules. Default logic as it stands does not *appropriately* resolve the lottery paradox.

What we have here, in fact, is the multiple extension problem all over again—but where it bites. For, as we saw, the default system can conclude that Tweety is a Hummingbird by ordering its consideration of default rules. By symmetrical reasoning we can have the default system infer any other alternative result. In other words, there is a different default extension available for each alternative default conclusion. As the differing content of these conflicting extensions describes what the default system remains inconclusive about, *no conclusion* about Tweety will be available. Etherington, et al. (1991) conclude in their recent discussion “since there is no basis for determining *which* assumptions to forego, however, any is as good as another . . . nothing can be assumed about the individual tickets” in the lottery (p. 225). Although we have seen that this kind of suspension of belief is wholly inadequate, it is hard to see how default logic alone can support any induction more substantial—it appears committed to endorsing Pollock’s Rule. It is trivial that *everything* is abnormal in *some* respect, otherwise it would not be a *single* thing; if those respects in which the objects under consideration differ are governed by default rules (as Poole 1991 argues is the normal case), then nothing short of direct observational knowledge of their properties will suffice to rid us of indecision. Default logic is impotent to carry us beyond direct, observational knowledge.

Some default logicians have suggested that such difficulties as we have encountered might be

avoided by imposing a preference order upon default rules. We have already seen an example of this in Touretzky's system. That preference order was based upon syntactic relations between the rules which are not satisfied by the bird-species rules, since they all have the same prerequisites. But preferences can be imposed upon the default rules regardless, and some might think that offers a way out of the arbitrariness brought on by the lottery paradox. However, there is no apparent basis for preferring one of the bird rules over another, any more than there is a basis for preferring one lottery ticket number over another. So the arbitrariness will remain, whether or not it is formalized in an explicit preference ranking for default rules.

Another version of the lottery for default logic is given by considering a single default rule as it applies over finitely many cases, rather than many rules applied to a single case, as above. Here too we have problems. Default inferences are default precisely because they have exceptions—that is, there is at least one member of the set of objects to which the default rule could be applied, but for which this would result in an error. Since the set of potential objects here will generally be finite, the situation for such a rule is isomorphic to a lottery with one or more winning tickets (exceptions). In other words, being stuck in the lottery paradox is the *normal* state of affairs, whether for the probabilist or for the default logician. So, the use of Pollock's Rule to suspend belief results in the suspension of virtually all inductive (default) inferential processes.

What does default logic have to say about the conjunction principle? That is, are the default conclusions arrived at indefinitely conjoinable? The definition of default proof given by Reiter (1980, p. 99) allows the free deployment of *any number* of default conclusions in building up the proof. Although this does not necessarily result in accepting their grand conjunction, it remains a form of conjoint reasoning. Most natural would be to give our default conclusions the same status as any other conclusion, making them available to all varieties of further reasoning. But granting that our default conclusions are *conclusions*, and worthy of residing in our corpora of beliefs, leaves the difficulty that the *only* constraint imposed is at the furthest reaches of such inference, when a known local inconsistency would otherwise arise: we have seen that the consistency requirement restrains the default logic system from careening straight into incoherence. That restraint is necessary, but does not go far enough—it ignores again the cumulative uncertainties of default conclusions. Defaults that are based upon other defaults cannot properly be regarded as safe as the original defaults. That is just to say that the warrant we have for accepting default conclusions is a matter of degree, it is a function of the kind and quantity of support for the default conclusions that can be brought to bear. But measuring *degrees* of support based upon the available evidence is a task that default logic ignores by design—in its restriction to purely symbolic, non-numeric means. In attempting to represent defeasible inference, default logicians have adopted a goal that is strictly incompatible with their logicist methods.

## 6.4 The Bayesian Alternative

Is all of this criticism of logicist approaches a tad unfair? What can quantitative methods offer to solve these problems?

Bayesian AI does not have generally accepted solutions to these problems.<sup>31</sup> What it does have is an approach to solving them which does not immediately degenerate into confusion. Given a probabilistic acceptance rule there is no requirement that the acceptances go so far as inconsistency, nor even—as with default logic—to inconsistency's door step. It is perfectly reasonable to impose a probabilistic constraint on the acceptance of assertions within a specific problem context. In the original lottery example, if our acceptance threshold was indeed 0.9, then such a constraint might stop us from accepting an individual  $\neg\Phi_i$  statement if the conjunction of



it with previously accepted statements about the lottery  $(\bigwedge_j \neg\Phi_j) \wedge \neg\Phi_i$  would have a probability below 0.9.

It will now be asked: if we are retaining such probabilistic information, why bother with qualitative acceptance? The answer must be, briefly, that with qualitative acceptance, we enable the deployment of qualitative inference, planning and decision making methods which have substantial computational advantages over their quantitative counterparts (cf. Harsanyi 1985). Nor are such advantages necessarily lost by imposing a probabilistic threshold upon the conjunction of statements accepted within a problem context—for heuristic means might be used for estimating their joint probability and avoiding a breach of the probabilistic threshold.

Another natural question is, How can the Bayesian system select one  $\neg\Phi_i$  to accept when it has nearly a million others—quantitatively identical—to choose from? If indeed there is *nothing* to distinguish one from the others, beyond its index, then I must agree with the suspension of belief. But the mere fact that there is no *probabilistic* distinction does not settle the matter. Typically there is a problem context to the inference—there will be one or more tickets of particular interest (e.g., mine). Or we may be concerned about some larger class of tickets collectively (e.g., rain vs. sun in the Sahara). The pragmatic concerns of dealing with a problem that suggests the deployment of qualitative reasoning resources must come into play before probabilistic acceptance can have any point.

As Etherington, et al. (1991) have complained, the fact that reasoning is *directed* by pragmatic concerns has been largely neglected in the construction of nonmonotonic formalisms. They go on to apply that point to the usual qualitative mechanisms for handling nonmonotonicity. By such means default logic and the other nonmonotonic formalisms can avoid the kinds of embarrassment we have looked at above, for if our concern is with *Tweety's* flying abilities, we need not be obliged to render simultaneously default verdicts on the flying abilities of all other birds. Thus, a properly *scoped* default logic need not be burdened with the peculiar tendencies of thought that both unrestrained default systems and unrestrained probabilistic systems suffered from.

Explicit consideration of pragmatics is then a way out of the lottery paradox for qualitative systems just as much as for Bayesian systems. There is good reason, however, for not following Etherington et al. in backing this fix to default logic. Restricting the scope of default reasoning to a particular set of objects and group of properties, while avoiding paradox, does not respond to the difficulty *underlying* the paradoxes—the accumulation of uncertainty in reasoning conjointly with multiple ampliative conclusions. In other words, the risk of error—something Etherington et al. recognize as relevant, to their credit (p. 231)—is sensitive to the depth and complexity of the conjoint reasoning, which may be left unconstrained by limiting its *scope alone*.<sup>32</sup> The application of a probabilistic threshold, or its heuristic counterpart, to such underlying conjunctions provides the missing sensitivity.

Etherington et al. conclude that for problems involving a broad scope probability theory is appropriate, while nonmonotonic logics “seem better suited to reasoning about small numbers of cases” (p. 232). If it be acknowledged that it is not the number of cases of interest but rather the number of inferences and their relative frailty, then this is surely a reasonable conclusion. Scoped and truncated default logic may be one of those qualitative methods that those of us constrained to suboptimality (i.e., all of us) must rely upon; and so default logic may find a useful place within the cognitive realm as one among many inferential mechanisms. But the scope of probability theory is nonetheless the entire range of inductive thought: it is only the low probability of error (or its low expected cost) that underwrites even such limited applications of default logics.

The most natural Bayesian alternative to scoped default reasoning is to replace default rules by conditional probabilities and to replace default inference by probabilistic acceptance. Bayesian AI can respond to the pressures of computational complexity by employing *scoped* deductive reasoning, or other qualitative reasoning, to the statements accepted. The scope involved will be determined in part by the pragmatics of the situation and in part by the operative probabilistic threshold (which itself must be sensitive to pragmatic issues, especially the expected cost of making an error in one's reasoning; cf. Kyburg 1965). Whether that threshold has been breached may be determined by examining the joint probability of what has been assumed or, if that probability is not readily available, by heuristic methods. The ideal result would be a hybrid reasoning system that uses probabilities where that is computationally feasible and acceptance-cum-qualitative methods otherwise. There must be at least three limitations imposed upon the acceptance rule for such an inference mechanism: (1) The pragmatic context must call for acceptance. If, for example, the problem is one of taking or not taking an explicit *bet*—a context which demands probabilistic reasoning—there is every reason *not* to detach a statement from its probability. (2) Within a particular problem context, the joint probability of accepted statements (whether determined heuristically or fully normatively) should stay above an appropriate threshold. (3) When, despite all precautions, inconsistency arises, the system must be capable of sophisticated belief revision, including the retraction of previously accepted statements.

## 7 Conclusion

Qualitative inference systems, if they are intended to apply to inductive problems, must address the lottery paradox, every bit as much as quantitative systems. The minimal requirement of weak consistency requires them to not just succumb to the paradox. But the goal of supporting general induction requires a rejection of the conjunction principle. Indeed, it is clear that none of our normative principles of *inference* can be read as applying without restriction or limit. Qualitative inference systems, to be viable candidates for modeling inductive inference, must then supply *reasoned* limits on conjunction (and the application of any other inference rules) that do justice to our concept of cognitive agency—that are, for example, not so severe as to render anything like the scientific inductions that we do impossible. However, qualitative methods do not appear to have the resources necessary for that task.

The reasons for constraining conjunction include the avoidance of local inconsistencies—which is a reason that qualitative systems can take advantage of, as default logic does. But we also saw in the case of default logic that such a constraint is insufficient. I propose instead that we adopt a hybrid approach: a model of a cognitive agent that allows the use of probabilistic information to support induction combined with qualitative methods of inference, including a *restricted* ability to conjoin prior inductive conclusions. Such a model has promise for avoiding both inductive paralysis in the face of uncertainty and computational paralysis in the face of complexity.

## Notes

1. Here I am following C.S. Peirce's account of ampli[f]iative inference (1940, pp. 180-181):

All our reasonings are of two kinds: 1. *Explicative, analytic, or deductive*; 2. *Ampliative, synthetic, or (loosely speaking) inductive*. In explicative reasoning, . . . .

But synthetic reasoning is of another kind. In this case the facts summed up in the conclusion are not among those stated in the premisses. They are different facts, as when one sees that the tide rises  $m$  times and concludes that it will rise the next time. These are the only inferences which increase our real knowledge, however useful the others may be.

I must point out that Ronald Loui (1991) has recently come out with a different account of ampliative inference, one which distinguishes inferences that use non-formalizable methods (“Ampliative inference is the result of rational nondeterministic nonmonotonic computation”; p. 153) rather than one which distinguishes inferences that encompass new information. Loui’s definition, however interesting, loses out on precedence and in any case fails to capture the more general concept of induction that I am interested in here.

2. But then *we* are endowed with a rich background in most problem contexts, so that we *need not* rely upon such simple forms of induction.

3. No doubt my definition of induction will continue to strike some as odd. Indeed, the definition of induction as inference from the particular to the general stems from Aristotle. Such forms of inference, however, do not do justice to the varieties of inferential methods to be found in scientific practice, according to the general (if not universal) consensus among philosophers of science in this century. Even Aristotle did not claim the sufficiency of enumerative induction, supplementing it with ‘intuitive inductions’ (cf. Kneale’s discussion of these points in 1949, pp. 24-48; or Black 1967). In the end, of course, it is a matter of choice what meaning we take for ‘induction’. But the differences between the various inductive methods (by my definition) are minor compared to those between induction and deduction, and so it is appropriate to emphasize this latter distinction.

4. Howson and Urbach (1989) provide an excellent survey of that literature, from an unabashedly Bayesian point of view. Glymour (1980) launched an influential assault upon Bayesianism; an anthology of responses is Earman (1983). Earman (1992) provides an intriguing discussion of the issues, which does not quite decide whether to be Bayesian or not. Glenn Shafer’s writings may be sampled to obtain another anti-Bayesian viewpoint (see, for example, Shafer 1985; a Bayesian response is Korb 1994). Finally, a recent attack on Bayesian conditioning is Bacchus, Kyburg, and Thalos (1990), responded to in Korb (1992b).

5. Here I join forces—in a limited way—with connectionists in opposing *exclusive* reliance upon qualitative, symbolic reasoning. Their objections have often been aimed at Newell and Simon’s Physical Symbol System Hypothesis. That hypothesis rules out the use of reasoning based upon real-valued variables (cf. their requirement (3), p. 116), and so is antagonistic to probabilism in principle (if not in practice, when limited-precision arithmetic renders the full range of reals no more accessible to the probabilist or connectionist than to the logicist).

6. By normative AI I mean research directed at the production of any artificially intelligent system, as opposed to descriptive AI which aims at modeling natural intelligence—what Simon calls computational psychology.

7. Although we can hardly expect floppies to be used for such copying, we have been given no reason to believe that—once we understand our brains—we would be unable to either alter existing brains or construct new ones having the information possessed by someone else (disregarding ethical considerations).

8. McCarthy himself does not confuse product and process. Rather, he asserts that (1968, p. 405)

... in order for a program to be capable of learning something it must first be capable

of being told it . . . . Once this is achieved, we may be able to tell the advice taker how to learn from experience.

The initial thought here is dubious; certainly the biological model suggests that learning can precede language (and therefore telling). But I am probably nit-picking: no doubt what McCarthy has in mind is that *representations* are required for learning, so we should first concentrate on developing adequate representations, which could then just be inserted into our computer system. But in any case there is no reason to be optimistic about McCarthy's next step. That the system has means to *represent* (describe, reason about, etc.) learning procedures does not mean that the system has means to *execute* such procedures. However, it is true that in Lisp (the AI language of McCarthy's invention) procedures and data have the same form of representation. Therefore, by telling the Advice Taker the right statement (i.e., procedure), the Advice Taker can "learn" how to learn. But this is just to require that we humans solve the problem of machine learning in the first place—by coming to know what procedure to feed into the Advice Taker—which is just another way of saying that the theorem-proving, rote-learning Advice Taker has nothing much to teach us about machine learning. I.e., the AI problem remains unsolved until we come up with the missing learning program.

9. While humans may operate in this way, it is not strictly necessary that contradictory statements be *accepted*—which is what we typically mean by an inconsistent system—but only that they be jointly contemplated by the reasoning system.

10. When I claim here that birds fly, I will normally intend this to be understood as asserting that individual birds within the domain of discourse are capable of flight. It is a reasonable presumption that there are contexts of discourse that render the inference in the text both reasonable and non-monotonic. Following a principle of charity regarding the claims of default logicians, I will not explore the many possible contexts of discourse where the inference would be either unreasonable or monotonic (such as interpreting "birds generally fly" as the universal "all birds fly").

11. Since there is no decision procedure for consistency in first-order logic, any practical implementation of default logic will have to rely upon a heuristic search for inconsistency in  $K \cup \{JUS(d)\}$ .

12. In this context the denial of  $A \rightarrow B$  means  $A \rightarrow \neg B$ , in contrast to the negation, which would be  $\neg (A \rightarrow B)$ .

13. Of course, 'Nixon is a Republican' is not a universal generalization, but to let the example go through we could replace 'Nixon' by some uniquely identifying description of Nixon and get a universal statement.

14. Intuitively, an extension  $K^*$  of  $K$  is any deductively closed (the intended meaning of the suffix '\*') superset of  $K$  that is consistent (if  $K$  was) and that has new content added only through default proof and that cannot be extended further using default rules (i.e., it is the smallest fixed point of the default implication operator).

15. It is worth noting, however, that any logic which treats universal generalizations and default rules alike (as does Pearl's default logic, Pearl 1989) will be unable to make use of the inconsistency above, and will be constrained, apparently, to decide the case as ambiguous.

16. If you do not like this silly story, then you may consider the following syllogism instead:

Most college students are adults  
Most adults are employed  
Most college students are employed

This is an example of "interacting defaults" (or, the problem of specificity) introduced to the

literature in Reiter and Criscuolo (1981).

17. My thanks to John Winnie for pointing this out.

18. I am relying on the presentation of their argument in Besnard (1989, pp. 83-84).

19. Examination of the figure reveals, as Chris Wallace has pointed out to me, that there is a far better rule available than  $\text{At-Meeting}(\text{John}) \rightarrow \neg \text{At-Meeting}(\text{Bob})$ , namely

$$\frac{\text{ : } \neg \text{At-Meeting}(\text{Bob})}{\neg \text{At-Meeting}(\text{Bob})}$$

i.e., we should *always* conclude that Bob is absent, if that conclusion is consistent with what we know. This example shows that high conditional probability is not a *sufficient* condition for (reason for) adopting a default rule: we want *useful* default rules and, in particular, default rules that do not serve as well as some competing rule ought not to be employed.

20. It should be noted that the reading of default rules as expressing typicality is the *ordinary* one in the literature of default and nonmonotonic logics, and is quite commonly thought to be incompatible with the probabilistic interpretation of default inference. (Cf. for example Reiter 1987, Etherington 1988, and Lukasiewicz 1990).

21. I do not know how Nutter comes down on the question of normative versus descriptive AI, so it is not clear that she would endorse any such anti-normative principle *as a guide to inference*. As a statement about the cognitive psychology of humans, typicality-based inference may be unobjectionable, but then it does not serve as an objection to probabilism within *normative* AI. Nutter's writings provide evidence of both naturalistic and normative tendencies. For a statement of the latter consider (1987, p. 374; my emphasis):

...designers of AI systems generally care less whether their systems "ought" to believe their answers than *how often those answers are right*. For systems whose judgements have practical consequences, we should measure and maximize that if we measure anything.

This seems to imply the opposite of the anti-normative principle I have attributed to her. And yet her bold assertion of the appropriateness of the default rule that birds fly under the hypothetical condition that we know most birds do not fly is as plain as day (1990, pp. 34-35). It may be that some convoluted consideration about normative versus descriptive contexts can make simultaneous sense of these disparate statements, but that is not my concern here. (And the possibly related point that the default rule does not express an assertion at all will be handled in the discussion of McCarthy's argument from linguistic conventions below.)

22. Of course, the situation is symmetric thus far. To the claim that typicality is a necessary ingredient for certain kinds of default rules I might offer a probabilistic counterexample. To manufacture one, if our environmental (non)policies have the effect of suddenly extinguishing all the birds other than kiwis (sorry, Tweety), then on probabilistic grounds we could accept the default rule  $\text{Bird} \rightarrow \neg \text{Flies}$ . But the sudden extinction would have little (immediate) effect on our understanding of 'bird' and would certainly not have changed how *we* had *previously* learned to think about birds. In other words, flying would remain prototypical. The default theorist might object that for that very reason the new default must be rejected. But the default theorist could hardly then make reference to the broader concern to minimize error—for minimizing error requires paying attention to probabilities.

23. Nutter (1990, p. 35) seems to think that this move *never* works, since species don't fly, only individuals do. However, when a normal person says that eagles are a flying species, we understand perfectly well that a new kind of flight—one reserved for *species*—is *not* being

entertained. (To be accurate eagles do not form a species; but calling them a species is as close as I will come to biological accuracy in this paper.)

24. Regarding fuzzy generalizations, perhaps it is worth noting that default logic also does not provide any analysis. In a default logic we can, of course, adopt a rule *Basketball-Player*  $\rightarrow$  *Tall*, but we can adopt a corresponding conditional probability as well, all without shedding light on fuzziness or claiming to have an adequate semantics in either case. There is an odd tendency in the literature to demand of Bayesians what is not demanded of others. For some discussion of the remaining cases of default generalizations, see chapter 5 of Korb (1992b).

25. This discussion of the lottery paradox reproduces material from Korb (1992a) by permission of the Philosophy of Science Association.

26. Of course, rejecting probabilistic acceptance is *not* tantamount to endorsing qualitative methods; that is, probabilistic acceptance has frequently been rejected by *probabilists*, most notably by many Bayesians.

27. Direct inference, in a somewhat more general form, is the sole source of probabilities in Kyburg's system (see his 1974) and is widely accepted as *one* source of probabilities among Bayesian philosophers as well.

28. Talk of rejecting and accepting events or subevents is meant here as a convenient shorthand for talk of rejecting and accepting propositions asserting the occurrence of those events.

29. Perhaps it is worth noting that normal statistical inference is also ruled out. For example, in order to have reason to reject the null hypothesis, it is necessary that we be able to assert prior to a statistical test that the outcome will not lie in the critical region on the assumption of the null hypothesis. But since we can partition the outcome space into regions each of which is as improbable as the critical region, Pollock's Rule obliges us to refrain from the normal statistical inference.

30. Perlis (1987), pp. 62ff., demonstrates the severe difficulties that the main nonmonotonic formalisms encounter in dealing with the lottery paradox, including circumscription and modal nonmonotonic logic.

31. Here I must confess that the criticisms I have adduced of qualitative approaches to defeasible inference do not leave only Bayesianism (with acceptance) whole—all quantitative means of measuring support remain unscathed. But Bayesianism's differences with the Dempster-Shafer calculus and fuzzy logic are beyond the reach of this discussion. Likewise, although I suggest reasons for acceptance, I cannot here explore my differences with that form of Bayesianism that eschews all talk of the acceptance of hypotheses, prominently defended by Richard Jeffrey.

32. Whether this is so depends upon just what is meant by limiting the scope of default logic. If the scope is constrained to a small set of objects of size  $k$  and a small number of binary properties, say  $n$ , then however deep the reasoning it can jointly assume only  $2^{n+k}$  propositions about them. But even with small  $n$  and  $k$  this point remains insufficient for supporting scoped default reasoning without consideration of the joint probability of what is being assumed, which will be sensitive for example to whatever probabilistic dependencies obtain between the different assumptions.

## References

- Bacchus, F., Kyburg, H., and Thalos, M. (1990) "Against Conditionalization," *Synthese* 85: 475-506.
- Besnard, P. (1989) *An Introduction to Default Logic*. Berlin: Springer Verlag.
- Black, M. (1967) "Induction," pp. 169-181 in volume 4 of P. Edwards (ed.) *The Encyclopedia*

of *Philosophy*. New York: Macmillan.

Carnap, R. (1962) *Logical Foundations of Probability*, second edition. University of Chicago.

Church, A. (1936) "An Unsolvable Problem of Number Theory," *American Journal of Mathematics* 58: 345-363.

Cohen, L.J. (1983) "Belief, Acceptance, and Probability," *Behavioral and Brain Sciences* 6: 248-249.

de Finetti, B. (1937) "Foresight: Its Logical Laws, Its Subjective Sources," *Annales de l'Institut Henri Poincaré*, vol. 7. Translated in H. Kyburg and H. Smokler (eds.) *Studies in Subjective Probability*, New York: Wiley, 1963.

Dubois, D., H. Farreny, and H. Prade (1985) "Sur divers problèmes inhérents à l'automatisation des raisonnements de sens commun," *Congrès AFCET-RFIA, Grenoble 1*: 321-328.

Earman, J. (ed.) (1983) *Testing Scientific Theories. Minnesota Studies in the Philosophy of Science*, vol. 10. University of Minnesota.

Earman, J. (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass: MIT.

Etherington, D.W. (1988) *Reasoning with Incomplete Information*. Los Altos, Calif.: Morgan Kaufmann.

Etherington, D.W., S. Kraus, D. Perlis (1991) "Nonmonotonicity and the Scope of Reasoning," *Artificial Intelligence* 52: 221-261.

Fahlman, J. (1979) *NETL: A System for Representing and Using Real-World Knowledge*. Cambridge: MIT.

Ginsberg, M. (ed) (1987) *Readings in Nonmonotonic Reasoning*. San Mateo: Morgan Kaufmann.

Glymour, C. (1980) *Theory and Evidence*. Princeton University.

Gödel, K. (1931) "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandten Systeme I," *Monatshefte für Mathematik und Physik* 38: 173-198.

Grice, P. (1975) "Logic and Conversation," in P. Cole and J.P. Morgan (eds.) *Syntax and Semantics, Volume 3: Speech Acts*.

Harman, G. (1965) "The Inference to the Best Explanation," *Philosophical Review* 70: 88-95.

Harnad, S. (1989) "Minds, Machines and Searle," *Journal of Experimental and Theoretical AI* 1: 5-25.

Harnad, S. (1990) "The Symbol Grounding Problem," *Physica D* 42: 335-346.

Harsanyi, J.C. (1985) "Acceptance of Empirical Statements: A Bayesian Theory without Cognitive Utilities," *Theory and Decision* 18: 1-30.

Hayes, P. (1977) "In Defense of Logic," *Proceedings IJCAI-77*, pp. 559-565.

Hayes, P. (1979) "The Naive Physics Manifesto," in D. Michie (ed.) *Expert Systems in the Micro-electronic Age*. Edinburgh: Edinburgh University.

Hayes, P. (1985) "The Second Naive Physics Manifesto," pp. 1-36 in Hobbs and Moore 1985.

Hayes-Roth, F., D.A. Waterman, and D. Lenat (1983) *Building Expert Systems*. Reading, Mass: Addison-Wesley.

Hintikka, J. and R. Hilpinen (1966) "Knowledge, Acceptance, and Inductive Logic," in Hintikka and Suppes (1966), pp. 1-20.

Hintikka, J. and P. Suppes (eds.) (1966) *Aspects of Inductive Logic*. Amsterdam: North Holland.

Hobbs, J.R. and R.C. Moore (eds.) (1985) *Formal Theories of the Commonsense World*. Norwood, N.J.: Ablex.

Howson, C. and P. Urbach (1989) *Scientific Reasoning: The Bayesian Approach*. Open Court.

- Israel, D. (1980) "What's Wrong with Non-Monotonic Logic?" *Proceedings AAAI-80*. Stanford, Calif.
- Israel, D. (1985) "A Short Companion to the Naive Physics Manifesto," pp. 427-447 in Hobbs and Moore 1985.
- Kneale, W. (1949) *Probability and Induction*. Oxford: Clarendon Press.
- Korb, K.B. (1992a) "The Collapse of Collective Defeat: Lessons from the Lottery Paradox," *PSA 1992*, vol. I.
- Korb, K.B. (1992b) *A Pragmatic Bayesian Platform for Automating Scientific Induction*. Ph.D. dissertation, Indiana University.
- Korb, K.B. (1994) "Infinitely Many Resolutions of Hempel's Paradox," pp. 138-149 in R. Fagin (ed.) *Theoretical Aspects of Reasoning about Knowledge, V*, Morgan Kaufmann.
- Kyburg, H. (1961) *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University.
- Kyburg, H. (1965) "Probability, Rationality and a Rule of Detachment," pp. 301-310 in Bar-Hillel (ed.) *Logic, Methodology and Philosophy of Science*. Amsterdam: North-Holland.
- Kyburg, H. (1970) "Conjunctivitis," in M. Swain (ed.) *Induction, Acceptance, and Rational Belief*. Dordrecht: D. Reidel.
- Kyburg, H. (1974) *The Logical Foundations of Statistical Inference*. Dordrecht: D. Reidel.
- Levi, I. (1983) "Conjunctive Bliss," *Behavioral and Brain Sciences 6*: 54-5.
- Lindley, D.V. (1971) *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics.
- Loui, R.P. (1991) "Ampliative Inference, Computation, and Dialectic," pp. 141-155 in R. Cummins and J. Pollock *Philosophy and AI: Essays at the Interface*, MIT.
- Lukasiewicz, W. (1990) *Non-monotonic Reasoning: Formalization of Commonsense Reasoning*. Chichester: Ellis Horwood.
- McCarthy, J. (1968) "Programs with Common Sense," pp. 403-418 in M. Minsky (ed.) *Semantic Information Processing*. Cambridge: MIT.
- McCarthy, J. (1980) "Circumscription—a Form of Non-monotonic Reasoning," *Artificial Intelligence 13*: 27-39.
- McCarthy, J. (1986) "Applications of Circumscription to Formalizing Common-sense Knowledge," *Artificial Intelligence 28*: 89-116.
- McDermott, D. (1987) "A Critique of Pure Reason," *Computational Intelligence 3*: 151-160.
- Minsky, M. (1974) *A Framework for Representing Knowledge*. AI Memo 306. Cambridge, Mass: MIT AI Lab.
- Newell, A., J.C. Shaw, and H. Simon (1957) "Empirical Explorations with the Logic Theory Machine," reprinted in E.A. Feigenbaum and J. Feldman (eds.) *Computers and Thought*. New York: McGraw-Hill, 1963.
- Newell, A. and H. Simon (1976) "Computer Science as Empirical Inquiry: Symbols and Search," *Communications of the ACM 19*: 113-126.
- Nutter, J.T. (1987) "Uncertainty and Probability," *Proceedings IJCAI-87*; pp. 373-379. Los Altos: Morgan Kaufmann.
- Nutter, J.T. (1990) "Knowledge Representation Issues in Default Reasoning," *Journal of Experimental and Theoretical AI 2*: 25-54.
- Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufmann.
- Pearl, J. (1989) "Probabilistic Semantics for Nonmonotonic Reasoning: A Survey," *Proceedings, First International Conference on Principles of Knowledge Representation and Reasoning*. Reprinted in Shafer and Pearl (1990), pp. 699-710.



- Peirce, C.S. (1940) "The Probability of Induction," in J. Buchler (ed.) *Philosophical Writings of Peirce*. Routledge and Kegan Paul.
- Perlis, D. (1987) "On the Consistency of Commonsense Reasoning," *Computational Intelligence 2*: 180-190.
- Pollock, J.L. (1987) "Defeasible Reasoning," *Cognitive Science 11*: 481-518.
- Pollock, J.L. (1989) *How To Build a Person: A Prolegomenon*. Cambridge: MIT.
- Poole, D. (1989) "Explanation and Prediction: An Architecture for Default and Abductive Reasoning," *Computational Intelligence 5*: 97-110.
- Poole, D. (1991) "The Effect of Knowledge on Belief: Conditioning, Specificity, and the Lottery Paradox in Default Reasoning," *Artificial Intelligence 49*: 281-307.
- Popper, K. (1959) *The Logic of Scientific Discovery*. New York: Basic Books. Translation, with new appendices, of *Logik der Forschung*, Vienna 1934.
- Ramsey, F.P. (1931) *The Foundations of Mathematics and Other Logical Essays*, edited by R.B. Braithwaite. New York: Humanities Press.
- Reiter, R. (1978) "On Reasoning by Default," *Theoretical Issues in Natural Language Processing 2*: 210-18.
- Reiter, R. (1980) "A Logic for Default Reasoning," *Artificial Intelligence 13*: 81-132.
- Reiter, R. (1987) "Nonmonotonic Reasoning," *Annual Review of Computer Science 2*: 147-186.
- Reiter, R. and G. Criscuolo. (1981) "On Interacting Defaults," *IJCAI 1981*, 270-76.
- Shafer, G. (1985) "Conditional Probability," *International Statistical Review 53*: 261-277.
- Shafer, G. and J. Pearl (eds.) (1990) *Readings in Uncertain Reasoning*. San Mateo: Morgan Kaufmann.
- Simon, H. (1983) "Why Should Machines Learn?" pp. 25-37 in R.S. Michalski, J.G. Carbonell and T.M. Mitchell (eds.) *Machine Learning*. Los Altos, Calif: Morgan Kaufmann.
- Stalnaker, R. (1984) *Inquiry*. Cambridge: MIT.
- Touretzky, D. (1984) "Implicit ordering of defaults in inheritance systems," *AAAI-84*. Reprinted in Ginsberg (1987).
- Turing, A.M. (1937) "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society 42*: 230-265.
- Wittgenstein, L. (1958) *Philosophical Investigations*, translated by G.E.M. Anscombe from the third German edition.