(This is a slightly revised version of "The theory of nomic probability", *Synthese* **90** (1992), 263-300. It summarizes much of the material published in *Nomic Probability and the Foundations of Induction*, Oxford, 1990. The paper was written before the book, but appeared after it.)

The Theory of Nomic Probability II

John L. Pollock
Department of Philosophy
University of Arizona
Tucson, Arizona 85721
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

1. Introduction

Probability theorists divide into two camps — the proponents of subjective probability, and the proponents of objective probability. Opinion has it that subjective probability has carried the day, but I think that such a judgment is premature. I have argued elsewhere that there are deep incoherencies in the notion of subjective probability. Accordingly, I find myself in the camp of objective probability. The consensus is, however, that the armies of objective probability are in disarray. The purpose of this paper is to outline a theory of objective probability that rectifies that. Such a theory must explain the meaning of objective probability, show how we can discover the values of objective probabilities, clarify their use in decision theory, and demonstrate how they can be used for epistemological purposes. The theory of nomic probability aims to do all that. I have presented most of the technical details of the theory elsewhere, but the bits and pieces are scattered about in a number of different journal articles. The purpose of this paper is to tie them all together and give the reader a philosophical overview of the theory without going into all of the logical and mathematical details.

There are two kinds of physical laws — statistical and nonstatistical. Statistical laws are probabilistic. I will call the kind of probability involved in statistical laws *nomic probability*. The best way to understand nomic probability is by looking first at non-statistical laws. What distinguishes such laws from material generalizations of the form "$(\forall x)(Fx \rightarrow Gx)$" is that they are not just about actual F's. They are about "all the F's there could be", that is, they are about "physically possible F's". I call non-statistical laws *nomic generalizations*. Nomic generalizations can be expressed in English using the locution "Any F would be a G". I will symbolize this nomic generalization as "$F \Rightarrow G$". It can be roughly paraphrased as telling us that any physically possible F would be G. *Physical possibility*, symbolized "$\Diamond_p Q$", means that Q is logically consistent with the set of all true nomic generalizations.

I propose that we think of nomic probabilities as analogous to nomic generalizations. Just as "$F \Rightarrow G$" tells that us any physically possible F would be G, for heuristic purposes we can think of the statistical law "$\mathbf{prob}(G/F) = r$" as telling us that the proportion of physically possible F's that would be G's is r. For instance, pretend it is a law of nature that at any given time, there are exactly as many electrons as protons. Then in every physically possible world, the proportion of electrons-or-protons that are electrons is 1/2. It is then reasonable to regard the probability of a particular particle being an electron given that it is either an electron or a proton as 1/2. Of course, in the general case, the proportion of G's that are F's will vary from one possible world to another. **prob**(F/G) then "averages" these proportions across all physically possible worlds. The mathematics of this averaging process is complex, and I will say more about it below.

Nomic probability is illustrated by any of a number of examples that are difficult for frequency theories. For instance, consider a physical description D of a coin, and suppose there is just one coin of that description and it is never flipped. On the basis of the description D together with our knowledge of physics we might conclude that a coin of this description is a fair coin, and hence the probability of a flip of a coin of description D landing heads is 1/2. In saying this we are not talking about relative frequencies — as there are no flips of coins of description D, the relative frequency does not exist. Or suppose instead that the single coin of description D is flipped just once, landing heads, and then destroyed. In that case the relative frequency is 1, but we would still insist that the probability of a coin of that description landing heads is 1/2. The reason for the difference between the relative frequency and the probability is that the probability statement is in some sense subjunctive or counterfactual. It is not just about actual flips, but about possible flips as well. In saying that the probability is 1/2, we are saying that out of all physically possible flips of coins of description D, 1/2 of them would land heads. To illustrate nomic probability with a more realistic example, in physics we often want to talk about the probability of some event in simplified circumstances that have never occurred. For example, the typical problem given students in a quantum mechanics class is of this character. The relative frequency does not exist, but the nomic probability does and that is what the students are calculating.

The theory of nomic probability will be a theory of probabilistic reasoning. I will not attempt to *define* 'nomic probability' in terms of simpler concepts, because I doubt that that can be done. If we have learned anything from twentieth century philosophy, it should be that philosophically interesting concepts are rarely definable. You cannot solve the problem of perception by defining 'red' in terms of 'looks red', you cannot solve the problem of other minds by defining 'person' in terms of behavior, and you cannot provide foundations for probabilistic reasoning by defining 'probability' in terms of relative frequencies. In general, the principles of reasoning involving various categories of concepts are primitive constituents of our conceptual framework and cannot be derived from definitions. The task of the epistemologist must simply be to state the principles precisely. That is my objective for probabilistic reasoning.

Probabilistic reasoning has three constituents. First, there must be rules prescribing how to ascertain the numerical values of nomic probabilities on the basis of observed relative frequencies. Second, there must be "computational" principles that enable us to infer the values of some nomic probabilities from others. Finally, there must be principles enabling us to use nomic probabilites to draw conclusions about other matters.

The first element of this account will consist largely of a theory of statistical induction. The second element will consist of a calculus of nomic probabilities. The final element will be an account of how conclusions not about nomic probabilities can be inferred from premises about nomic probability. This will have two parts. First, it seems clear that under some circumstances, knowing that certain probabilities are high can justify us in holding related non-probabilistic beliefs. For example, I know it to be highly probable that the date appearing on a newspaper is the correct date of its publication. (I do not know that this is always the case — typographical errors do occur). On this basis, I can arrive at a justified belief regarding today's date. The epistemic rules describing when high probability can justify belief are called *acceptance rules*. The acceptance rules endorsed by the theory of nomic probability will constitute the principal novelty of that theory. The other fundamental principles that will be adopted as primitive assumptions about nomic probability are all of a computational nature. They concern the logical and mathematical structure of nomic probability, and in effect amount to nothing more than an elaboration of the standard probability calculus. It is the acceptance rules that give the theory its unique flavor and comprise the main epistemological machinery making it run.

It is important to be able to make another kind of inference from nomic probabilities. We can make a distinction between "definite" probabilities and "indefinite" probabilities. A definite probability is the probability that a particular proposition is true. Indefinite probabilities, on the other hand, concern properties rather than propositions. For example, the probability of a smoker getting cancer is not about any particular smoker. Rather, it relates the property of being a smoker and the property of getting cancer. Nomic probabilities are indefinite probabilities. This is automatically the case for any probabilities derived by induction from relative frequencies, because relative frequencies relate properties. But for many practical purposes, the probabilities we are really interested in are definite probabilities. We want to know how probable it is that it will rain today, that Bluenose will win the third race, that Sally will have a heart attack, etc. It is probabilities of this sort that are involved in practical reasoning. Thus the first three elements of our analysis must be augmented by a fourth element. That is a theory telling us how to get from indefinite probabilities to definite probabilities. We judge that there is a twenty percent probability of rain today, because the indefinite probability of its raining in similar circumstances is believed to be about .2. We think it unlikely that Bluenose will win the third race because he has never finished above seventh in his life. We judge that Sally is more likely than her sister to have a heart attack because Sally smokes like a furnace and drinks like a fish, while her sister is a nun who jogs and lifts weights. We take these facts about Sally and her sister to be relevant because we know that they affect the indefinite probability of a person having a heart attack. That is, the indefinite probability of a person who smokes and drinks having a heart attack is much greater than the indefinite probability for a person who does not smoke or drink and is in good physical condition. Inferences from indefinite probabilities to definite probabilities are called *direct inferences*. A satisfactory theory of nomic probability must include an account of direct inference.

To summarize, the theory of nomic probability will consist of (1) a theory of statistical induction, (2) an account of the computational principles allowing some probabilities to be derived from others, (3) an account of acceptance rules, and (4) a theory of direct inference.

2. Computational Principles

It might seem that the calculus of nomic probabilities should just be the classical probability calculus. But this overlooks the fact that nomic probabilities are indefinite probabilities. Indefinite probabilities operate on properties, including relational properties of arbitrarily many places. This introduces logical relationships into the theory of nomic probability that are ignored in the classical probability calculus. One simple example is the "principle of individuals":

(IND) $\mathbf{prob}(Axy\ /\ Rxy\ \&\ y{=}b) = \mathbf{prob}(Axb/Rxb)$.

This is an essentially relational principle and is not a theorem of the classical probability calculus. It might be wondered how there can be general truths regarding nomic probability that are not theorems of the classical probability calculus. The explanation is that, historically, the probability calculus was devised with definite probabilities in mind. The standard versions of the probability calculus originate with Kolmogoroff [1933] and are concerned with "events". The relationship between the calculus of indefinite probabilities and the calculus of definite probabilities is a bit like the relationship between the predicate calculus and the propositional calculus. Specifically, there are principles regarding relations and quantifiers that must be added to the classical probability calculus to obtain a reasonable calculus of nomic probabilities.

In developing the calculus of nomic probabilities, I propose that we make further use of the heuristic model of nomic probability as measuring proportions among physically possible objects. The statistical law "**prob**$(G/F) = r$" can be regarded as telling us that the proportion of physically possible F's that would be G is r. Treating probabilities in terms of proportions proves to be a useful approach for investigating the logical and mathematical structure of nomic probability.

Proportions operate on sets. Given any two sets A and B we can talk about the proportion of members of B that are also members of A. I will symbolize "the proportion of members of B that are in A" as "$\rho(A/B)$". The concept of a proportion is a general measure-theoretic notion. The theory of proportions is developed in detail in Pollock [1987a], and I will say more about it below. But first consider how we can use it to derive computational principles governing nomic probability. The derivation is accomplished by making more precise our explanation of nomic probability as measuring proportions among physically possible objects. Where F and G are properties and G is not counterlegal (*i*.e, it is physically possible for there to be G's), we can regard **prob**(F/G) as the proportion of physically possible G's that would be F's. This suggests that we define:

(2.1) If $\Diamond_p(\exists x)Gx$ then **prob**$(F/G) = \rho(\mathrm{F}/\mathrm{G})$

where F is the set of all physically possible F's and G is the set of all physically possible G's. This forces us to consider more carefully just what we mean by 'a physically possible F'. We cannot just mean 'a possible object that is F in some physically possible world', because the same object can be F in one physically possible world and non-F in another. Instead, I propose to understand a physically possible F to be an ordered pair $\langle w,x\rangle$ such that w is a physically possible world (i.e., one having the same physical laws as the actual world) and x is an F at w. We then define:

(2.2) $\mathrm{F} = \{\langle w,x\rangle \mid w$ is a physically possible world and x is F at $w\}$;
$\mathrm{G} = \{\langle w,x\rangle \mid w$ is a physically possible world and x is G at $w\}$.

With this understanding, we can regard nomic probabilities straightforwardly as in (2.1) as measuring proportions between sets of physically possible objects. (2.1) must be extended to include the case of counterlegal probabilities, but I will not go into that here.

(2.1) reduces nomic probabilities to proportions among sets of physically possible objects. The next task is to investigate the theory of proportions. That investigation is carried out in Pollock [1987a] and [1990] and generates a calculus of proportions that in turn generates a calculus of nomic probabilities. The simplest and least problematic talk of proportions concerns finite sets. In that case proportions are just frequencies. Taking $\#X$ to be the cardinality of a set X, relative frequencies are defined as follows:

(2.3) If X and Y are finite and Y is nonempty then $\mathrm{freq}[X/Y] = \#(X\cap Y)/\#Y$.

We then have the *Frequency Principle*:

(2.4) If X and Y are finite and Y is nonempty then $\rho(X/Y) = \mathrm{freq}[X/Y]$.

But we also want to talk about proportions among infinite sets. The concept of a proportion in such a case is an extension of the concept of a frequency. The simplest laws governing proportions are those contained in the classical probability calculus, which can be axiomatized as follows:

(2.5) $0 \leq \rho(X/Y) \leq 1$

(2.6) If $Y \subseteq X$ then $\rho(X/Y) = 1$.

(2.7) If $Z \neq \varnothing$ and $Z \cap X \cap Y = \varnothing$ then $\rho(X \cup Y/Z) = \rho(X/Z) + \rho(Y/Z)$.

(2.8) $\rho(X \cap Y/Z) = \rho(X/Z) \cdot \rho(Y/X \cap Z)$.

Given the theory of proportions and the characterization of nomic probabilities in terms of proportions of physically possible objects, we can derive a powerful calculus of nomic probabilities. Much of the theory is rather standard looking. For example, the following versions of the standard axioms for the probability calculus follow from (2.5)–(2.8):

(2.9) $0 \leq \mathbf{prob}(F/G) \leq 1$.

(2.10) If $(G \Rightarrow F)$ then $\mathbf{prob}(F/G) = 1$.

(2.11) If $\Diamond$H and $[H \Rightarrow \sim(F \& G)]$ then $\mathbf{prob}(F \vee G/H) = \mathbf{prob}(F/H) + \mathbf{prob}(G/H)$.

(2.12) If $\Diamond_p(\exists x)$Hx then $\mathbf{prob}(F \& G/H) = \mathbf{prob}(F/H) \cdot \mathbf{prob}(G/F \& H)$.

The theory of proportions resulting from (2.4) – (2.8) might be termed "the Boolean theory of proportions", because it is only concerned with the Boolean operations on sets. In this respect, it is analogous to the propositional calculus. However, in $\rho(X/Y)$, X and Y might be sets of ordered pairs, i.e., relations. There are a number of principles that ought to hold in that case but are not contained in the Boolean theory of proportions. The classical probability calculus takes no notice of relations, and to that extent it is seriously inadequate. For example, the following *Cross Product Principle* would seem to be true:

(2.13) If $C \neq \varnothing$, $D \neq \varnothing$, $A \subseteq C$, and $B \subseteq D$, then $\rho(A \times B/C \times D) = \rho(A/C) \cdot \rho(B/D)$.

In the special case in which A, B, C and D are finite, the cross product principle follows from the frequency principle, but in general it is not a consequence of the classical probability calculus.

I have found that experienced probability theorists tend to raise two sorts of spurious objections at this point. The first is that the cross product principle is not a new principle — "It can be found in every text on probability theory under the heading of 'product spaces"'. That is quite true, but irrelevant to the point I am making. My point is simply that this is a true principle regarding proportions that is not a theorem of the classical probability calculus. The second spurious objection acknowledges that the cross product principle is not a theorem of the probability calculus but goes on to insist that that is as it should be because the principle is false. It is "explained" that the cross product does not hold in general because it assumes the statistical independence of the members of C and D. This objection is based upon a confusion, and it is important to get clear on this confusion because it will affect one's entire understanding of the theory of proportions. The confusion consists of not distinguishing between probabilities and proportions. These are two quite different things. What the probability theorist is thinking is that we should not endorse the following principle regarding *probabilities*:

$$\mathbf{prob}(Ax\&By \,/\, Cx\&Dy) = \mathbf{prob}(Ax \,/\, Cx)\cdot\mathbf{prob}(Bx \,/\, Dx)$$

because the C's and the D's need not be independent of one another. For example, if $A = B$ and $C = D$ this principle would entail that

$$\mathbf{prob}(Ax\&Ay \,/\, Cx\&Cy) = (\mathbf{prob}(A/C))^2.$$

But suppose it is a law that $(Cx\&Cy) \Rightarrow (Ax \equiv Ay)$. Then we should have

$$\mathbf{prob}(Ax\&Ay \,/\, Cx\&Cy) = \mathbf{prob}(A/C),$$

and hence we would obtain the absurd result that $\mathbf{prob}(A/C) = (\mathbf{prob}(A/C))^2$. But all of this pertains to probabilities — not proportions. The cross product principle for proportions does not imply the cross product principle for probabilities. Proportions are simply relative measures of the sizes of sets. If we consider the case in which $A = B$ and $C = D$, what the cross product principle tells us is that the relative measure of $A\times A$ is the square of the relative measure of A, i.e.,

$$\rho(A\times A/C\times C) = (\rho(A/C))^2,$$

and this principle is undeniable. For example, when A and C are finite this principle is an immediate consequence of the fact that if A has n members then $A\times A$ has n^2 members. Talk of independence makes no sense when we are talking about proportions.

The fact that the cross product principle is not a consequence of the classical probability calculus demonstrates that the probability calculus must be strengthened by the addition of some "relational" axioms in order to axiomatize the general theory of proportions. The details of the choice of relational axioms turn out to be rather complicated, and I will not pursue them further here. However, the theory developed in Pollock [1987a] and [1990] turns out to have some important consequences. One of these concerns probabilities of probabilities. On many theories, there are difficulties making sense of probabilities of probabilities, but there are no such problems within the theory of nomic probability. '**prob**' can relate any two properties, including properties defined in terms of nomic probabilities. Our theorem is:

(PPROB) If r is a rigid designator of a real number and $\Diamond[(\exists x)Gx \;\&\; \mathbf{prob}(F/G) = r]$ then $\mathbf{prob}\big(F \,/\, G \;\&\; \mathbf{prob}(F/G) = r\big) = r$.

The most important theorem of the calculus of nomic probabilities is the *Principle of Agreement*, which I will now explain. This theorem follows from an analogous principle regarding proportions, and I will begin by explaining that principle. First note a rather surprising combinatorial fact (at least, surprising to the uninitiated in probability theory). Consider the proportion of members of a finite set B that are in some subset A of B. Subsets of B need not exhibit the same proportion of A's, but it is a striking fact of set theory that subsets of B *tend* to exhibit *approximately* the same proportion of A's as B, and both the strength of the tendency and the degree of approximation improve as the size of B increases. More precisely, where "$x \approx_\varepsilon y$" means "x is approximately equal to y, the difference being at most ε", the following is a theorem of set theory:

(2.14) For every $\varepsilon,\delta > 0$, there is an n such that if B is a finite set containing at least n members then $\text{freq}\big[\text{freq}[A/X] \approx_\delta \text{freq}[A/B] \ / \ X \subseteq B\big] > 1\text{-}\varepsilon$.

It seems inescapable that when B becomes infinite, the proportion of subsets agreeing with B to any given degree of approximation should become 1. This is *The Principle of Agreement for Proportions*:

(2.15) If B is infinite and $\rho(A/B) = p$ then for every $\varepsilon > 0$, $\rho\big(\rho(A/X) \approx_\varepsilon p \ / \ X \subseteq B\big) = 1$.

This principle seems to me to be undeniable. It is simply a generalization of (2.14) to infinite sets. It is shown in Pollock [1987a] that this principle can be derived within a sufficiently strong theory of proportions.

The importance of the Principle of Agreement For Proportions is that it implies a Principle of Agreement For Nomic Probabilities. Let us say that H is a *subproperty* of G iff H nomically implies G and H is not counterlegal:

(2.16) $H \preceq G$ iff $\Diamond_p \exists H$ & $\Box_p \forall(H \rightarrow G)$.

Strict subproperties are subproperties that are restricted to the set of physically possible worlds:

(9.5) $H \prec G$ iff (1) H is a subproperty of G and (2) if $\langle w, x_1, \ldots, x_n \rangle \in H$ then w is a physically possible world.

In effect, strict subproperties result from chopping off those parts of properties that pertain to physically impossible worlds. The following turns out to be an easy consequence of the Principle of Agreement for Proportions:

(AGREE) If F and G are properties and there are infinitely many physically possible G's and $\textbf{prob}(F/G) = p$ (where p is a nomically rigid designator) then for every $\varepsilon > 0$, $\textbf{prob}\big(\textbf{prob}(F/X) \approx_\varepsilon p \ / \ X \prec G\big) = 1$.

This is *The Principle of Agreement for Probabilities*. (AGREE) is the single most important computational principle of nomic probability. It lies at the heart of the theory of direct inference, and that makes it fundamental to the theory of statistical and enumerative induction.

A profligate ontology of sets of possible worlds and possible objects underlies the constructions I have been describing. It deserves to be emphasized that this use of proportions and sets of physically possible objects is just a way of getting the mathematical structure of nomic probability right. It will play no further role in the theory. Accordingly, if ontological scruples prevent one from accepting this metaphysical foundation for nomic probability, one can instead view it as a merely heuristic model for arriving at an appropriate formal mathematical structure. Once we have adopted that structure we can forget about possible worlds and possible objects. This formal mathematical structure will be coupled with epistemological principles that are ontologically neutral, and the theory of nomic probability will be derived from that joint basis.

3. The Statistical Syllogism

Rules telling us when it is rational to believe something on the basis of high probability are called *acceptance rules*. The philosophical literature contains numerous proposals for acceptance rules, but most proceed in terms of definite probabilities rather than indefinite probabilities. There is, however, an obvious candidate for an acceptance rule that proceeds in terms of nomic probability. This is the *Statistical Syllogism*, whose traditional formulation is something like the following:

> Most *A*'s are *B*'s.
> This is an *A*.
> ________________
> Therefore, this is a *B*.

It seems clear that we often reason in roughly this way. For instance, on what basis do I believe what I read in the newspaper? Certainly not that everything printed in the newspaper is true. No one believes that. But I do believe that *most* of what is published in the newspaper is true, and that justifies me in believing individual newspaper reports. Similarly, I do not believe that every time a piece of chalk is dropped, it falls to the ground. Various circumstances can prevent that. It might be snatched in midair by a wayward hawk, or suspended in air by Brownian movement, or hoisted aloft by a sudden wind. None of these are at all likely, but they are possible. Consequently, all I can be confident of is that chalk, when dropped, will almost always fall to the ground. Nevertheless, when I drop a particular piece of chalk, I expect it to fall to the ground.

"Most *A*'s are *B*'s" can have different interpretations. It may mean simply that most actual *A*'s are *B*'s. But sometimes it is used to talk about more than just actual *A*'s. For instance, in asserting that most pieces of chalk will fall to the ground when dropped, I am not confining the scope of 'most' to actual pieces of chalk. If by some statistical fluke it happened that at this instant there was just one piece of chalk in the world, all others having just been used up and no new ones manufactured for the next few seconds, and that single remaining piece of chalk was in the process of being dropped and snatched in midair by a bird, I would not regard that as falsifying my assertion about most pieces of chalk.

At least sometimes, 'most' statements can be cashed out as statements of nomic probability. On that construal, "Most *A*'s are *B*'s" means "$\mathbf{prob}(B/A)$ is high". I think this is the most natural construal of my claim about the behavior of pieces of chalk. This suggests the following acceptance rule, which can be regarded as a more precise version of the statistical syllogism:

> $\mathbf{prob}(B/A)$ is high.
> Ac
> ________________
> Therefore, Bc.

Clearly, the conclusion of the statistical syllogism does not follow deductively from the premises. Furthermore, although the premises may often make it reasonable to accept the conclusion, that is not always the case. For instance, I may know that most ravens are black, and Josey is a raven, but I may also know that Josey is an albino raven and hence is not black. The premises of the statistical syllogism can at most create a presumption in favor of the conclusion, and that presumption can be defeated by contrary information. In other words, the inference licensed by this rule must be a *defeasible* inference: The inference is a reasonable one in the absence of conflicting information, but it is possible to have conflicting information in the face of which the inference becomes unreasonable.

In general, if P is a defeasible reason for Q, there can be two kinds of defeaters for P. *Rebutting defeaters* are reasons for denying Q in the face of P. To be contrasted with rebutting defeaters are *undercutting defeaters*, which attack the connection between the defeasible reason and its conclusion rather than attacking the conclusion itself. For example, something's looking red to me is a reason for me to think it that it is red. But if I know that x is illuminated by red lights and such illumination often makes things look red when they are not, then I cannot be justified in believing that x is red on the basis of its looking red to me. What I know about the illumination constitutes an undercutting defeater for my defeasible reason. An undercutting defeater for P as a defeasible reason for believing Q is a reason for denying that P would not be true unless Q were true. To illustrate, knowing about the peculiar illumination gives me a reason for denying that x would not look red to me unless it were red. The technical details of reasoning with defeasible reasons and defeaters are discussed in Pollock [1987] and [1995].

As a first approximation, the statistical syllogism can be formulated as follows:

(3.1) If $r > 0.5$ then "Ac and $\mathbf{prob}(B/A) \geq r$" is a defeasible reason for "Bc", the strength of the reason depending upon the value of r.

It is illuminating to consider how this rule handles the lottery paradox. Suppose you hold one ticket in a fair lottery consisting of one million tickets, and suppose it is known that one and only one ticket will win. Observing that the probability is only .000001 of a ticket being drawn given that it is a ticket in the lottery, it seems reasonable to accept the conclusion that your ticket will not win. But by the same reasoning, it will be reasonable to believe, for each ticket, that it will not win. However, these conclusions conflict jointly with something else we are justified in believing, namely, that some ticket will win. Assuming that we cannot be justified in believing each member of an explicitly contradictory set of propositions, it follows that we are not warranted in believing of each ticket that it will not win. But this is no problem for our rule of statistical syllogism as long as it provides only a defeasible reason. What is happening in the lottery paradox is that the defeasible reason is defeated.

The lottery paradox is a case in which we have defeasible reasons for a number of conclusions but they collectively defeat one another. This illustrates the *principle of collective defeat*. This principle will turn out to be of considerable importance in probability theory, so I will say a bit more about it. Starting from propositions we are objectively justified in believing, we may be able to construct arguments supporting some further propositions. But that does not automatically make those further propositions warranted, because some propositions supported in that way may be defeaters for steps of some of the other arguments. That is what happens in cases of collective defeat. Suppose we are warranted in believing some proposition R and we have equally good defeasible reasons for each of $P_1,...,P_n$, where $\{P_1,...,P_n\}$ is a minimal set of propositions deductively inconsistent with R (i.e., it is a set deductively inconsistent with R and has no proper subset that is deductively inconsistent with R). Then for each i, the conjunction "R & P_1 & ... & P_{i-1} & P_{i+1} & ... & P_n" entails $\sim P_i$. Thus by combining this entailment with the arguments for R and $P_1,...,P_{i-1},P_{i+1},...,P_n$ we obtain an argument for $\sim P_i$ that is as good as the argument for P_i. Thus we have equally strong support for both P_i and $\sim P_i$, and hence we could not reasonably believe either on this basis, i.e., neither is warranted. This holds for each i, so none of the P_i is warranted. They collectively defeat one another. Thus the simplest version of the principle of collective defeat can be formulated as follows:

(3.2) If we are warranted in believing R and we have equally good independent defeasible reasons for each member of a minimal set of propositions deductively inconsistent with R, and none of these defeasible reasons is defeated in any other way, then none of the propositions in the set is warranted on the basis of these defeasible reasons.

Although the principle of collective defeat allows the principle (3.1) of statistical syllogism to escape the lottery paradox, it turns out that the very fact that (3.1) can handle the lottery paradox in this way shows that it cannot be correct. The difficulty is that every case of high probability can be recast in a form that makes it similar to the lottery paradox. The details of this are a bit complicated, and they are spelled out in Pollock [1983a] and [1990], so I will not repeat them here. The difficulty can be traced to the assumption that A and B in (3.1) can be arbitrary formulas. Basically, we need a constraint to rule out arbitrary disjunctions. It turns out that disjunctions create repeated difficulties throughout the theory of probabilistic reasoning. This is easily illustrated in the case of (3.1). For instance, it is a theorem of the probability calculus that $\mathbf{prob}(F/G\vee H) \geq \mathbf{prob}(F/G)\cdot\mathbf{prob}(G/G\vee H)$. Consequently, if $\mathbf{prob}(F/G)$ and $\mathbf{prob}(G/G\vee H)$ are sufficiently large, it follows that $\mathbf{prob}(F/G\vee H) \geq r$. For example, because the vast majority of birds can fly and because there are many more birds than giant sea tortoises, it follows that most things that are either birds or giant sea tortoises can fly. If Herman is a giant sea tortoise, (3.1) would give us a reason for thinking that Herman can fly, but notice that this is based simply on the fact that most birds can fly, which should be irrelevant to whether Herman can fly. This indicates that arbitrary disjunctions cannot be substituted for B in (3.1).

Nor can arbitrary disjunctions be substituted for A in (3.1). By the probability calculus, $\mathbf{prob}(F\vee G/H) \geq \mathbf{prob}(F/H)$. Therefore, if $\mathbf{prob}(F/H)$ is high, so is $\mathbf{prob}(F\vee G/H)$. Thus, because most birds can fly, it is also true that most birds can either fly or swim the English Channel. By (3.1), this should be a reason for thinking that a starling with a broken wing can swim the English Channel, but obviously it is not.

There must be restrictions on the properties A and B in (3.1). To have a convenient label, let us say that B is *projectible with respect to A* iff (3.1) holds. What we have seen is that projectibility is not closed under disjunction, i.e., neither of the following hold:

If C is projectible with respect to both A and B, then C is projectible with respect to $(A\vee B)$.

If A and B are both projectible with respect to C, then $(A\vee B)$ is projectible with respect to C.

On the other hand, it seems fairly clear that projectibility is closed under conjunction.

It seems that in formulating the principle of statistical syllogism, we must build in an explicit projectibility constraint:

(A1) If F is projectible with respect to G and $r > .5$, then "Gc & $\mathbf{prob}(F/G) \geq r$" is a defeasible reason for believing "Fc", the strength of the reason depending upon the value of r.

Of course, if we define projectibility in terms of (3.1), (A1) becomes a mere tautology, but the intended interpretation of (A1) is that *there is* a relation of projectibility between properties, holding in important cases, such that "Gc & $\mathbf{prob}(F/G) \geq r$" is a defeasible reason for "Fc" when F is projectible with respect to G. To have a fully adequate theory we must augment (A1) with an account of projectibility, but that proves very difficult and I have no account to propose. At best, our conclusions about closure conditions provide a partial account. Because projectibility is closed under conjunction but not under disjunction, it follows that it is not closed under negation. Similar considerations establish that it is not closed under the formation of conditionals or biconditionals. It is not clear how it behaves with quantifiers. Although projectibility is not closed under negation, it seems likely that negations of "simple" projectible properties are projectible. For instance, both 'red' and 'nonred' are projectible with respect to 'robin'. A reasonable hypothesis is that there is a large class P containing most properties that are intuitively "logically simple", and whenever A and B are conjunctions of members of P and negations of members of P, A is projectible with respect to B. This is at best a sufficient condition for projectibility, however, because we will find numerous cases of projectibility involving properties that are logically more complex than this.

The term 'projectible' comes from the literature on induction. Goodman [1955] was the first to observe that principles of induction require a projectibility constraint. I have deliberately chosen the term 'projectible' in formulating the constraint on (A1). This is because in the theory of nomic probability, principles of induction become theorems rather than primitive postulates. The acceptance rules provide the epistemological machinery that make the theory run, and the projectibility constraint in induction turns out to derive from the projectibility constraint on (A1). It is the same notion of projectibility that is involved in both cases.

The reason provided by (A1) is only a defeasible reason. As with any defeasible reason, it can be defeated by having a reason for denying the conclusion. The reason for denying the conclusion constitutes a rebutting defeater. But there is also an important kind of undercutting defeater for (A1). In (A1), we infer the truth of "Fc" on the basis of probabilities conditional on a limited set of facts about c (i.e., the facts expressed by "Gc"). But if we know additional facts about c that alter the probability, that defeats the defeasible reason:

(D1) If F is projectible with respect to H then "Hc & $\mathbf{prob}(F/G\&H) \neq \mathbf{prob}(F/G)$" is an undercutting defeater for (A1).

I will refer to these as *subproperty defeaters*. (D1) amounts to a kind of "total evidence requirement". It requires us to make our inference on the basis of the most comprehensive facts regarding which we know the requisite probabilities.

(A1) is not the only defensible acceptance rule. There is another acceptance rule that is related to (A1) rather like modus tollens is related to modus ponens:

(A2) If F is projectible with respect to G then "$\sim Fc$ & $\mathbf{prob}(F/G) \geq r$" is a defeasible reason for "$\sim Gc$", the strength of the reason depending upon the value of r.

(A2) is easily illustrated. For example, on the basis of quantum mechanics, we can calculate that it is highly probable that an energetic electron will be deflected if it passes within a certain distance of a uranium atom. We observe that a particular electron was not deflected, and so conclude that it did not pass within the critical distance. Reasoning in this way with regard to the electrons used in a scattering experiment, we arrive at conclusions about the diameter of a uranium atom.

It seems clear that (A1) and (A2) are closely related. I suggest that they are consequences of a single stronger principle:

(A3) If F is projectible with respect to G then "$\mathbf{prob}(F/G) \geq r$" is a defeasible reason for the conditional "$Gc \rightarrow Fc$", the strength of the reason depending upon the value of r.

(A1) can then be replaced by an instance of (A3) and modus ponens, and (A2) by an instance of (A3) and modus tollens. Accordingly, I will regard (A3) as the fundamental probabilistic acceptance rule. Just as in the case of (A1), when we use (A3) we are making an inference on the basis of a limited set of facts about *c*. That inference should be defeated if the probability can be changed by taking more facts into account. This indicates that the defeater for (A2) and (A3) should be the same as for (A1):

(*D*) If *F* is projectible with respect to (*G*&*H*) then "*Hc* & **prob**(*F*/*G*&*H*) ≠ **prob**(*F*/*G*)" is an undercutting defeater for (A1), (A2), and (A3).

I take it that (A3) is actually quite an intuitive acceptance rule. It amounts to a rule saying that, when *F* is projectible with respect to *G*, if we know that most *G*'s are *F*, that gives us a reason for thinking of any particular object that it is an *F* if it is a *G*. The only surprising feature of this rule is the projectibility constraint. (A3) is the basic epistemic principle from which all the rest of the theory of nomic probability is derived.

4. Direct Inference and Definite Probabilities

Nomic probability is a species of indefinite probability, but as I remarked above, for many purposes we are more interested in definite ("single case") probabilities. In particular, this is required for decision-theoretic purposes. The probabilities required for decision theory must have a strong epistemic element. For example, any decision on what odds to accept on a bet that Bluenose will win the next race must be based in part on what we know about Bluenose, and when our knowledge about Bluenose changes so will the odds we are willing to accept. Such probabilities are mixed physical/epistemic probabilities.

What I call 'classical direct inference' aims to derive physical/epistemic definite probabilities from indefinite probabilities. The basic idea behind classical direct inference was first articulated by Hans Reichenbach: in determining the probability that an individual *c* has a property *F*, we find the narrowest reference class *X* for which we have reliable statistics and then infer that **PROB**(*Fc*) = **prob**(*Fx*/*x*∈ *X*). For example, insurance rates are calculated in this way. There is almost universal agreement that direct inference is based upon some such principle as this, although there is little agreement about the precise form the theory should take. In Pollock [1983] and [1984], I argued that classical direct inference should be regarded as proceeding in accordance with two epistemic rules. Let us say that a proposition is warranted for a cognizer just in case further reasoning from his current epistemic situation could put him in a position where he is justified in believing the proposition and in which no additional reasoning would make him unjustified. Let "**W**φ" abbreviate "φ is warranted". Then the two rules are:

(4.1) If *F* is projectible with respect to *G* then "**prob**(*F*/*G*) = *r* & **W**(*Gc*) & **W**(*P* ↔ *Fc*)" is a defeasible reason for "**PROB**(*P*) = *r*".

(4.2) If *F* is projectible with respect to *H* then "**prob**(*F*/*H*) ≠ **prob**(*F*/*G*) & **W**(*Hc*) & *H* ⪯ *G*" is an undercutting defeater for (4.1).

Principle (4.2) formulates a kind of subproperty defeat for direct inference, because it says that probabilities based upon more specific information take precedence over those based upon less specific information. Note the projectibility constraint in these rules. That constraint is required to avoid various paradoxes of direct inference that turn in an essential way on disjunctions.

To illustrate this account of direct inference, suppose we know that Herman is a 40 year old resident of the United States who smokes. Suppose we also know that the probability of a 40 year old resident of the United States having lung cancer is 0.1, but the probability of a 40 year old smoker who resides in the United States having lung cancer is 0.3. If we know nothing else that is relevant we will infer that the probability of Herman having lung cancer is 0.3. (4.1) provides us with one defeasible reason for inferring that the probability is 0.1 and a second defeasible reason for inferring that the probability is 0.3. However, the latter defeasible reason is based upon more specific information, and so by (4.2) it takes precedence, defeating the first defeasible reason and leaving us justified in inferring that the probability is 0.3.

I believe that (4.1) and (4.2) are correct rules of classical direct inference, but I also believe that the nature of direct inference has been fundamentally misunderstood. Direct inference is taken to govern inferences from indefinite probabilities to definite probabilities, but it is my contention that such "classical" direct inference rests upon parallel inferences from indefinite probabilities to indefinite probabilities. The basic rule of classical direct inference is that if F is projectible with respect to G and we know "$\mathbf{prob}(F/G) = r$ & $\mathbf{W}(Gc)$" but do not know anything else about c that is relevant, this gives us a reason to believe that $\mathsf{PROB}(Fc) = r$. Typically, we will know c to have other projectible properties H but not know anything about the value of $\mathbf{prob}(F/G\&H)$ and so be unable to use the latter in direct inference. But if the direct inference from "$\mathbf{prob}(F/G) = r$" to "$\mathsf{PROB}(Fc) = r$" is to be reasonable, there must be a presumption to the effect that $\mathbf{prob}(F/G\&H) = r$. If there were no such presumption then we would have to regard it as virtually certain that $\mathbf{prob}(F/G\&H) \neq r$ (after all, there are infinitely many possible values that $\mathbf{prob}(F/G\&H)$ could have), and so virtually certain that there is a true subproperty defeater for the direct inference. This would make the direct inference to "$\mathsf{PROB}(Fc) = r$" unreasonable. Thus classical direct inference presupposes the following principle regarding indefinite probabilities:

(4.3) If F is projectible with respect to G then "$H \preceq G$ & $\mathbf{prob}(F/G) = r$" is a defeasible reason for "$\mathbf{prob}(F/H) = r$".

Inferences in accord with (4.3) comprise *non-classical direct inference*. (4.3) amounts to a kind of principle of insufficient reason, telling us that if we have no reason to think otherwise, it is reasonable for us to anticipate that conjoining H to G will not affect the probability of F.

A common reaction to (4.3) is that it is absurd — perhaps trivially inconsistent. This reaction arises from the observation that in a large number of cases, (4.3) will provide us with defeasible reasons for conflicting inferences or even defeasible reasons for inferences to logically impossible conclusions. For example, since in a standard deck of cards a spade is necessarily black and the probability of a black card being a club is one half, (4.3) gives us a defeasible reason to conclude that the probability of a spade being a club is one half, which is absurd. But this betrays an insensitivity to the functioning of defeasible reasons. A defeasible reason for an absurd conclusion is automatically defeated by the considerations that lead us to regard the conclusion as absurd. Similarly, defeasible reasons for conflicting conclusions defeat one another. If P is a defeasible reason for Q and R is a defeasible reason for $\sim Q$, then P and R rebut one another and both defeasible inferences are defeated. No inconsistency results. That this sort of case occurs with some frequency in non-classical direct inference should not be surprising, because it also occurs with some frequency in classical direct inference. In classical direct inference we very often find ourselves in the position of knowing that c has two logically independent properties G and H, where $\mathbf{prob}(F/G) \neq \mathbf{prob}(F/H)$. When that happens, classical direct inferences from these two probabilities conflict with one another, and so each defeasible reason is a defeater for the other, with the result that we are left without an undefeated direct inference to make.

Although (4.3) is not trivially absurd, it is not self-evidently true either. The only defense I have given for it so far is that it is required for the legitimacy of inferences we commonly make. That is a reason for thinking that it is true, but we would like to know why it is true. The answer to this question is supplied by the principle of agreement and the acceptance rule (A1). We have the following instance of (A1):

(4.4) If "$\mathbf{prob}(F/X) \approx_\delta r$" is projectible with respect to "$X \preceq G$" then "$H \preceq G$ & $\mathbf{prob}\big(\mathbf{prob}(F/X) \approx_\delta r \,/\, X \preceq G\big) = 1$" is a defeasible reason for "$\mathbf{prob}(F/H) \approx_\delta r$".

If we assume that the property "$\mathbf{prob}(F/X) \approx_\delta r$" is projectible with respect to "$X \preceq G$" whenever F is projectible with respect to G, then it follows that:

(4.5) If F is projectible with respect to G then "$H \preceq G$ & $\mathbf{prob}\big(\mathbf{prob}(F/X) \approx_\delta r \,/\, X \preceq G\big) = 1$" is a defeasible reason for "$\mathbf{prob}(F/H) \approx_\delta r$".

By the principle of agreement, for each $\delta > 0$, "$\mathbf{prob}(F/G) = r$" entails "$\mathbf{prob}(\mathbf{prob}(F/X) \approx_\delta r \,/\, X \preceq G) = 1$", so it follows that:

(4.6) If F is projectible with respect to G then for each $\delta > 0$, "$H \preceq G$ & $\mathbf{prob}(F/G) = r$" is a defeasible reason for "$\mathbf{prob}(F/H) \approx_\delta r$".

Although there are some differences of detail, this is strikingly similar to the principle (4.3) of nonclassical direct inference, and I showed in Pollock [1984] that (4.3) follows from this.

Similar reasoning enables us to derive the following defeater for (4.3):

(4.7) If F is projectible with respect to J then "$H \preceq J$ & $J \preceq G$ & $\mathbf{prob}(F/J) \neq \mathbf{prob}(F/G)$" is an undercutting defeater for (4.3).

I will refer to these as *subproperty defeaters* for nonclassical direct inference.

We now have two kinds of direct inference — classical and non-classical. Direct inference has traditionally been identified with classical direct inference, but I believe that it is most fundamentally non-classical direct inference. The details of classical direct inference are all reflected in non-classical direct inference. If we could identify definite probabilities with certain indefinite probabilities, we could derive the theory of classical direct inference from the theory of non-classical direct inference. This can be done by noting that the following is a theorem of the calculus of nomic probabilities:

(4.8) If $\Box(Q \leftrightarrow Sa_1...a_n)$ and $\Box(Q \leftrightarrow Bb_1...b_m)$ and $\Box[Q \rightarrow (P \leftrightarrow Ra_1...a_n)]$ and $\Box[Q \rightarrow (P \leftrightarrow Ab_1...b_m)]$, then $\mathbf{prob}(Rx_1...x_n \,/\, Sx_1...x_n \;\&\; x_1{=}a_1 \;\&\; ... \;\&\; x_n{=}a_n) = \mathbf{prob}(Ay_1...y_m \,/\, By_1...y_m \;\&\; y_1 = b_1 \;\&\; ... \;\&\; y_m = b_m)$.

This allows us to define a kind of definite probability as follows:

(4.9) prob$(P/Q) = r$ iff for some n, there are n-place properties R and S and objects $a_1,...,a_n$ such that $\Box(Q \leftrightarrow Sa_1...a_n)$ and $\Box[Q \rightarrow (P \leftrightarrow Ra_1...a_n)]$ and **prob**$(Rx_1...x_n \ / \ Sx_1...x_n \ \& \ x_1 = a_1 \ \& \ ... \ \& \ x_n = a_n) = r$.

prob(P/Q) is an *objective* definite probability. It reflects the state of the world, not the state of our knowledge. The definite probabilities at which we arrive by classical direct inference are not those defined by (4.9). However, if we let W be the conjunction of all warranted propositions, we can define a mixed physical/epistemic probability as follows:

(4.10) **PROB**(P) = prob(P/W)

(4.11) **PROB**(P/Q) = prob$(P/Q\&\mathrm{W})$.

Given this reduction of definite probabilities to indefinite probabilities, it becomes possible to *derive* principles (4.1) and (4.2) of classical direct inference from our principles of non-classical direct inference, and hence indirectly from (A3) and the calculus of nomic probabilities. The upshot of all this is that the theory of direct inference, both classical and nonclassical, consists of a sequence of *theorems* in the theory of nomic probability. We require no new assumptions in order to get direct inference. At the same time, we have made clear sense of the mixed physical/epistemic probabilities that are needed for decision theory.

5. Induction

The values of some nomic probabilities are derivable from the values of others using the calculus of nomic probabilities or the theory of nonclassical direct inference, but our initial knowledge of nomic probabilities must result from empirical observation of the world around us. This is accomplished by *statistical induction*. We observe the relative frequency of *F*'s in a sample of *G*'s, and then infer that **prob**(F/G) is approximately equal to that relative frequency. One of the main strengths of the theory of nomic probability is that precise principles of induction can be derived from (and hence justified on the basis of) the acceptance rules and computational principles we have already endorsed. This leads to a solution of sorts to the problem of induction.

Principles of statistical induction are principles telling us how to estimate probabilities on the basis of observed relative frequencies in finite samples. The problem of constructing such principles is sometimes called 'the problem of inverse inference'. All theories of inverse inference are similar in certain respects. In particular, they all make use of some form of Bernoulli's theorem. In its standard formulation, Bernoulli's theorem tells us that if we have n objects $b_1,...,b_n$ and for each i, **PROB**$(Ab_i) = p$, and any of these objects being A is statistically independent of which others of them are A, then the probability is high that the relative frequency of A's among $b_1,...,b_n$ is approximately p, and the probability increases and the degree of approximation improves as n is made larger. These probabilities can be computed quite simply by noting that on the stated assumption of independence, it follows from the probability calculus that

$$\textbf{PROB}(Ab_1 \,\&\, ... \,\&\, Ab_r \,\&\, {\sim}Ab_{r+1} \,\&\, ... \,\&\, {\sim}Ab_n)$$

$$= \textbf{PROB}(Ab_1)\cdot...\cdot\textbf{PROB}(Ab_r)\cdot\textbf{PROB}({\sim}Ab_{r+1})\cdot...\cdot\textbf{PROB}({\sim}Ab_n)$$

$$= p^r(1\text{-}p)^{n\text{-}r}.$$

There are $n!/r!(n\text{-}r)!$ distinct ways of assigning A-hood among $b_1,...,b_n$ such that freq$[A/\{b_1,...,b_n\}] = r/n$, so it follows by the probability calculus that

$$\textbf{PROB}(\text{freq}[A/\{b_1,...,b_n\}] = r/n) = n!p^r(1\text{-}p)^{n\text{-}r}/r!(n\text{-}r)!.$$

The right side of this equation is the formula for the binomial distribution. An interval $[p\text{-}\varepsilon,p+\varepsilon]$ around p will contain just finitely many fractions r/n with denominator n, so we can calculate the probability that the relative frequency has any one of those values, and then the probability of the relative frequency being in the interval is the sum of those probabilities.

Thus far everything is uncontroversial. The problem is what to do with the probabilities resulting from Bernoulli's theorem. Most theories of inverse inference, including most of the theories embodied in contemporary statistical theory, can be regarded as variants of a single intuitive argument that goes as follows. Suppose $\mathbf{prob}(A/B) = p$, and all we know about $b_1,...,b_n$ is that they are B's. Then by classical direct inference we can infer that for each i, $\textbf{PROB}(Ab_i) = p$. If the b_i's seem intuitively unrelated to one another then it seems reasonable to suppose they are statistically independent and so we can use Bernoulli's theorem and conclude that it is extremely probable that the observed relative frequency r/n lies in a small interval $[p\text{-}\varepsilon,p+\varepsilon]$ around p. This entails conversely that p is within ε of r/n, i.e., p is in the interval $[(r/n)\text{-}\varepsilon,(r/n)+\varepsilon]$. This becomes our estimate of p.

This general line of reasoning seems plausible until we try to fill in the details. Then it begins to fall apart. There are basically two problems. The first is the assumption of statistical independence that is required for the calculation involved in Bernoulli's theorem. That is a probabilistic assumption. Made precise, it is the assumption that for each Boolean conjunction B of the conjuncts "Ab_j" for $j \neq i$, $\textbf{PROB}(Ab_i/B) = \textbf{PROB}(Ab_i)$. It seems that to know this we must already have probabilistic information, but then we seem to be forced into an infinite regress. In practice, it is supposed that if the b_i's seem to "have nothing to do with one another" then they are independent in this sense, but it is hard to see how that can be justified noncircularly. We might try to solve this problem by adopting some sort of fundamental postulate allowing us to assume independence unless we have some reason for thinking otherwise. In a sense, I think that is the right way to go, but it is terribly *ad hoc*. It will turn out below that it is possible to replace such a fundamental postulate by a derived principle following from the parts of the theory of nomic probability that have already been established.

A much deeper problem for the intuitive argument concerns what to do with the conclusion that it is very probable that the observed frequency is within ε of p. It is tempting to suppose we can use our acceptance rule (A3) and reason:

If $\mathbf{prob}(A/B) = p$ then $\textbf{PROB}(\text{freq}[A/\{b_1,...,b_n\}] \in [p\text{-}\varepsilon,p+\varepsilon])$ is approximately equal to 1

so

if $\mathbf{prob}(A/B) = p$ then freq$[A/\{b_1,...,b_n\}] \in [p\text{-}\varepsilon,p+\varepsilon]$.

The latter entails

If freq$[A/\{b_1,...,b_n\}] = r/n$ then **prob**$(A/B) \in [(r/n)\text{-}\varepsilon,(r/n)+\varepsilon]$.

A rather shallow difficulty for this reasoning is that it is an incorrect use of (A3). (A3) concerns indefinite probabilities, while Bernoulli's theorem supplies us with definite probabilities. But let us waive that difficulty for the moment, because there is a much more profound difficulty. This is that the probabilities we obtain in this way have the structure of the lottery paradox. Given any point q in the interval $[p\text{-}\varepsilon,p+\varepsilon]$, we can find a small interval ε around it such that if we let I_q be the union of two intervals $[0,q\text{-}\varepsilon]\cup[q+\varepsilon,1]$, the probability of freq$[A/\{b_1,...,b_n\}]$ being in I_q is as great as the probability of its being in $[p\text{-}\varepsilon,p+\varepsilon]$. This is diagrammed in figure 1. The probability of the frequency falling in any interval is represented by the area under the curve corresponding to that interval. The curve is reflected about the x axis so that the probability for the interval $[p\text{-}\varepsilon,p+\varepsilon]$ can be represented above the axis and the probability for the interval I_q represented below the axis.

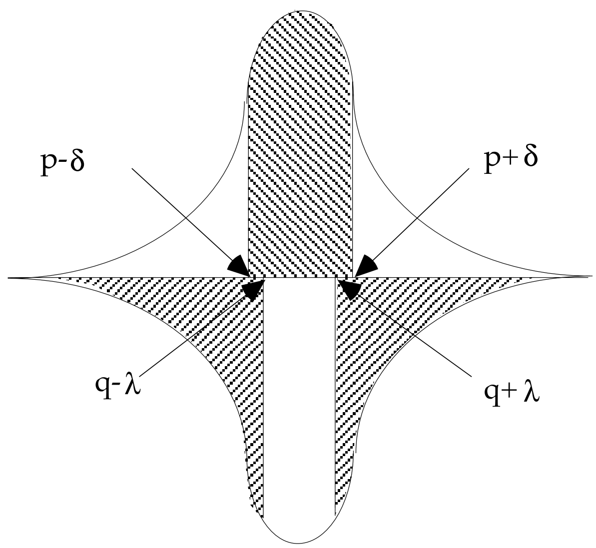

Figure 1. Intervals on the tails with the same probability as an interval at the maximum.

Next notice that we can construct a finite set $q_1,...,q_k$ of points in $[p\text{-}\varepsilon,p+\varepsilon]$ such that the "gaps" in the I_{q_i} collectively cover $[p\text{-}\varepsilon,p+\varepsilon]$. This is diagrammed in figure 2. For each $i \leq$ k, we have as good a reason for believing that freq$[A/\{b_1,...,b_n\}]$ is in I_{q_i} as we do for thinking it is in $[p\text{-}\varepsilon,p+\varepsilon]$, but these conclusions are jointly inconsistent. This is analogous to the lottery paradox. We have a case of collective defeat and thus are unjustified in concluding that the relative frequency is in the interval $[p\text{-}\varepsilon,p+\varepsilon]$.

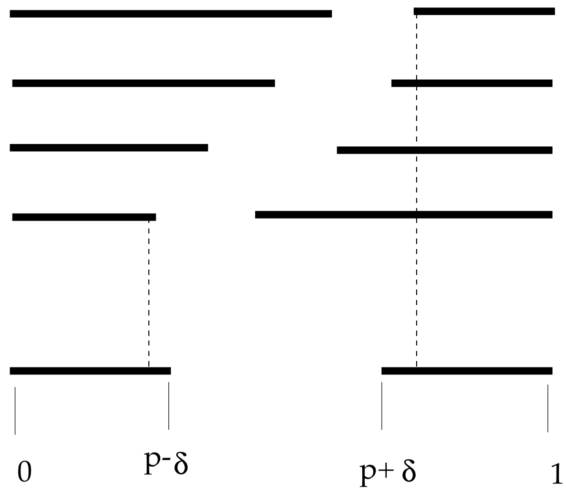

Figure 2. A collection of small gaps covers a large gap.

The intuitive response to the "lottery objection" consists of noting that [p-ε,p+ε] is an interval while the I_q are not. Somehow, it seems right to make an inference regarding intervals when it is not right to make the analogous inference regarding "gappy" sets. That is the line taken in orthodox statistical inference when confidence intervals are employed. But it is very hard to see why this should be the case, and some heavy duty argument is needed here to justify the whole procedure.

In sum, when we try to make the intuitive argument precise, it becomes apparent that it contains major lacunae. This does not constitute an utter condemnation of the intuitive argument. Because it is so intuitive, it would be surprising if it were not at least approximately right. Existing statistical theories tend to be *ad hoc* jury rigged affairs without adequate foundations, but it seems there must be some sound intuitions that statisticians are trying to capture with these theories. The problem is to turn the intuitive argument into a rigorous and defensible argument. That, in effect, is what my account of statistical induction does. The argument will undergo three kinds of repairs, creating what I call *the statistical induction argument*. First, it will be reformulated in terms of indefinite probabilities, thus enabling us to make legitimate use of our acceptance rules. Second, it will be shown that the gap concerning statistical independence can be filled by nonclassical direct inference. Third, the final step of the argument will be scrapped and replaced by a more complex argument not subject to the lottery paradox. This more complex argument will employ a principle akin to the *Likelihood Principle* of classical statistical inference.

The details of the statistical induction argument are complicated, and can be found in full in Pollock [1984a] and [1990]. I will try to convey the gist of the argument by focusing on a special case. Normally, **prob**(A/B) can have any value from 0 to 1. The argument is complicated by the fact that there are infinitely many possible values. Let us suppose instead that we somehow know that **prob**(A/B) has one of a finite set of values $p_1,...,p_k$. If we have observed a sample $X = \{b_1,...,b_n\}$ of B's and noted that only $b_1,...,b_r$ are A's (where A and ~A are projectible with respect to B), then the relative frequency freq[A/X] of A's in X is r/n. From this we want to infer that **prob**(A/B) is approximately r/n. Our reasoning proceeds in two stages, the first stage employing the theory of nonclassical direct inference, and the second stage employing our acceptance rules.

Stage I

Let us abbreviate "$x_1,...,x_n$ are distinct & Bx_1 & ... & Bx_n & **prob**$(Ax/Bx) = p$" as "θ_p". When $r \le n$, we have by the probability calculus:

(5.1) $\textbf{prob}(Ax_1 \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,/\, \theta_p)$
$= \textbf{prob}(Ax_1 \,/\, Ax_2 \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p)$
$\cdot\, ... \,\cdot\, \textbf{prob}(Ax_r \,/\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p)$
$\cdot\, \textbf{prob}({\sim}Ax_{r+1} \,/\, {\sim}Ax_{r+2} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p) \cdot ... \cdot \textbf{prob}({\sim}Ax_n \,/\, \theta_p)$.

Making θ_p explicit:

(5.2) $\textbf{prob}(Ax_i \,/\, Ax_{i+1} \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p)$
$= \textbf{prob}\big(Ax_i \,/\, x_1,...,x_n$ are distinct $\&\, Bx_1 \,\&\, ... \,\&\, Bx_n \,\&\, Ax_{i+1} \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ...$
$\&\, {\sim}Ax_n \,\&\, \textbf{prob}(A/B) = p\big)$.

Projectibility is closed under conjunction, so "Ax_i" is projectible with respect to "Bx_1 & ... & Bx_n & $x_1,...,x_n$ are distinct & Ax_{i+1} & ... & Ax_r & ${\sim}Ax_{r+1}$ & ... & ${\sim}Ax_n$". Given principles we have already endorsed it can be proven that whenever "Ax_i" is projectible with respect to "Fx" is projectible, it is also projectible with respect to "Fx & **prob**$(Ax/Bx) = p$". Consequently, "Ax_i" is projectible with respect to the reference property of (5.2). Thus a non-classical direct inference gives us a reason for believing that

$$\textbf{prob}(Ax_i \,/\, Ax_{i+1} \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p) = \textbf{prob}\big(Ax_i \,/\, Bx_i \,\&\, \textbf{prob}(A/B) = p\big),$$

which by principle (PPROB) equals p. Similarly, non-classical direct inference gives us a reason for believing that if $r < i \le n$ then

(5.3) $\textbf{prob}({\sim}Ax_i \,/\, {\sim}Ax_{i+1} \,\&\, ... \,\&\, {\sim}Ax_n \,\&\, \theta_p) = 1\text{-}p$.

Then from (5.1) we have:

(5.4) $\textbf{prob}(Ax_1 \,\&\, ... \,\&\, Ax_r \,\&\, {\sim}Ax_{r+1} \,\&\, ... \,\&\, {\sim}Ax_n \,/\, \theta_p) = p^r(1\text{-}p)^{n-r}$.

"freq[A / $\{x_1,...,x_n\}$] = r/n" is equivalent to a disjunction of $n!/r!(n\text{-}r)!$ pairwise incompatible disjuncts of the form "Ax_1 & ... & Ax_r & ${\sim}Ax_{r+1}$ & ... & ${\sim}Ax_n$", so by the probability calculus:

(5.5) $\textbf{prob}\big(\text{freq}[A \,/\, \{x_1,...,x_n\}] = r/n \,/\, \theta_p\big) = n!p^r(1\text{-}p)^{n-r}/r!(n\text{-}r)!$.

This is the formula for the binomial distribution.

This completes stage I of the statistical induction argument. This stage reconstructs the first half of the intuitive argument described above. Note that it differs from that argument in that it proceeds in terms of indefinite probabilities rather than definite probabilities, and it avoids having to make unwarranted assumptions about independence by using nonclassical direct inference instead. In effect, nonclassical direct inference gives us a reason for expecting independence unless we have evidence to the contrary. All of this is a consequence of our acceptance rule and computational principles.

Stage II

The second half of the intuitive argument ran afoul of the lottery paradox and seems to me to be irreparable. I propose to replace it with an argument using (A2). I assume at this point that if A is a projectible property then "freq[A/X] $\neq r/n$" is a projectible property of X. Thus the following conditional probability, derived from (5.5), satisfies the projectibility constraint of our acceptance rules:

(5.6) $\mathbf{prob}\big(\text{freq}[A \,/\, \{x_1,\ldots,x_n\}] \neq r/n \;/\; \theta_p\big) = 1 - \; n!p^r(1\text{-}p)^{n\text{-}r}/r!(n\text{-}r)!$.

Let $\mathbf{b}(n,r,p) = n!p^r(1\text{-}p)^{n\text{-}r}/r!(n\text{-}r)!$. For sizable n, $\mathbf{b}(n,r,p)$ is almost always quite small. E.g., $\mathbf{b}(50,20,.5) = .04$. Thus by (A2) and (5.6), for each choice of p we have a defeasible reason for believing that $\langle b_1,\ldots,b_n\rangle$ does not satisfy θ_p, i.e., for believing

$$\sim(b_1,\ldots,b_n \text{ are distinct } \& \; Bb_1 \;\&\; \ldots \;\&\; Bb_n \;\&\; \mathbf{prob}(A/B) = p).$$

As we know that "$b_1,\ldots,b_n$ are distinct & Bb_1 & ... & Bb_n" is true, this gives us a defeasible reason for believing that $\mathbf{prob}(A/B) \neq p$. But we know that for some one of $p_1,\ldots,p_k$, $\mathbf{prob}(A/B) = p_i$. This case is much like the case of the lottery paradox. For each i we have a defeasible reason for believing that $\mathbf{prob}(A/B) \neq p_i$, but we also have a counterargument for the conclusion that $\mathbf{prob}(A/B) = p_i$, viz:

$\mathbf{prob}(A/B) \neq p_1$
$\mathbf{prob}(A/B) \neq p_2$
.
.
.
$\mathbf{prob}(A/B) \neq p_{i\text{-}1}$
$\mathbf{prob}(A/B) \neq p_{i+1}$
.
.
.
$\mathbf{prob}(A/B) \neq p_k$
For some j between 1 and k, $\mathbf{prob}(A/B) = p_j$.
Therefore, $\mathbf{prob}(A/B) = p_i$.

But there is an important difference between this case and a fair lottery. For each i, we have a defeasible reason for believing that $\mathbf{prob}(A/B) \neq p_i$, but these reasons are not all of the same strength because the probabilities assigned by (5.6) differ for the different p_i's. The counterargument is only as good as its weakest link, so for some of the p_i's, the counterargument may not be strong enough to defeat the defeasible reason for believing that $\mathbf{prob}(A/B) \neq p_i$. This will result in there being a subset R (the *rejection class*) of $\{p_1,\ldots,p_k\}$ such that we can conclude that that for each $p \in$ R, $\mathbf{prob}(A/B) \neq p$, and hence $\mathbf{prob}(A/B) \notin$ R. Let A (the *acceptance class*) be $\{p_1,\ldots,p_k\}$ – R. It follows that we are justified in believing that $\mathbf{prob}(A/B) \in$ A. A will consist of those p_i's closest in value to r/n. Thus we can think of A as an interval around the observed frequency, and we are justified in believing that $\mathbf{prob}(A/B)$ lies in that interval.

To fill in some details, let us abbreviate 'r/n' as 'f', and make the simplifying assumption that for some i, $p_i = f$. $\mathbf{b}(n,r,p)$ will always be highest for this value of p, which means that (5.6) provides us with a weaker reason for believing that $\mathbf{prob}(A/B) \neq f$ than it does for believing that $\mathbf{prob}(A/B) \neq p_j$ for any of the other p_j's. It follows that f cannot be in the rejection class, because each step of the counterargument is better than the reason for believing that $\mathbf{prob}(A/B) \neq f$. On the other hand, "$\mathbf{prob}(A/B) \neq f$" will be the weakest step of the counterargument for every other p_j. Thus what determines whether p_j is in the rejection class is simply the comparison of $\mathbf{b}(n,r,p_j)$ to $\mathbf{b}(n,r,f)$. A convenient way to encode this comparison is by considering the ratio

(5.7) $\quad \mathrm{L}(n,r,p) = \mathbf{b}(n,r,p)/\mathbf{b}(n,r,f) = (p/f)^{\mathrm{nf}} \cdot ((1-p)/1-f)^{\mathrm{n}(1-f)}$.

$\mathrm{L}(n,r,p)$ is the *likelihood ratio* of "$\mathbf{prob}(A/B) = p$" to "$\mathbf{prob}(A/B) = f$". The smaller the likelihood ratio, the stronger is our on-balance reason for believing (despite the counterargument) that $\mathbf{prob}(A/B) \neq p$, and hence the more justified we are in believing that $\mathbf{prob}(A/B) \neq p$. Each degree of justification corresponds to a minimal likelihood ratio, so we can take the likelihood ratio to be a measure of the degree of justification. For each likelihood ratio α we obtain the α-*rejection class* R_α and the α-*acceptance class* A_α:

(5.8) $\quad \mathrm{R}_\alpha = \{p_i \mid \mathrm{L}(n,r,p_i) \leq \alpha\}$

(5.9) $\quad \mathrm{A}_\alpha = \{p_i \mid \mathrm{L}(n,r,p_i) > \alpha\}$.

We are justified to degree α in rejecting the members of R_α, and hence we are justified to degree α in believing that $\mathbf{prob}(A/B)$ is a member of A_α. If we plot the likelihood ratios, we get a bell curve centered around r/n, with the result that A_α is an interval around r/n and R_α consists of the tails of the bell curve. This is shown in figure 3. In interpreting this curve, remember that low likelihood ratios correspond to a high degree of justification for *rejecting* that value for $\mathbf{prob}(A/B)$, and so the region around r/n consists of those values we cannot reject, i.e., it consists of those values that might be the actual value.

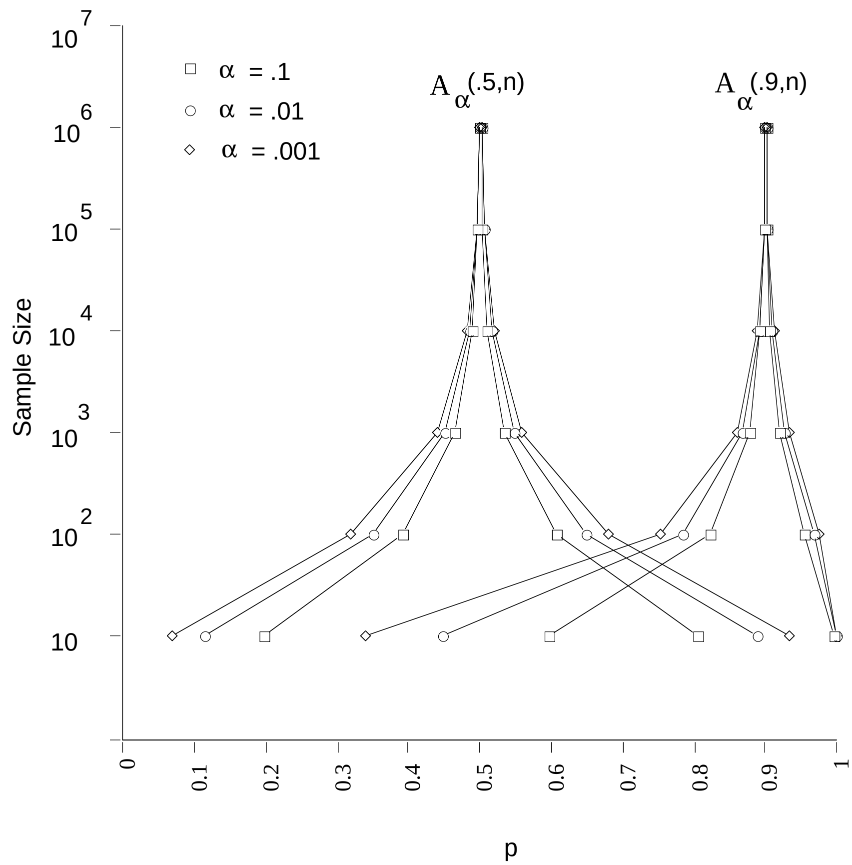

Figure 3. Acceptance Intervals

I have been discussing the idealized case in which we know that **prob**(*A*/*B*) has one of a finite set of values, but the argument can be generalized to apply to the continuous case as well. The argument provides us with justification for believing that **prob**(*A*/*B*) lies in a precisely defined interval around the observed relative frequency, the width of the interval being a function of the degree of justification. For illustration, some typical values of the acceptance interval are listed in table 1. Reference to the acceptance level reflects the fact that attributions of warrant are indexical. Sometimes an acceptance level of .1 may be reasonable, at other times an acceptance level of .01 may be required, and so forth.

Table 1. Values of $A_\alpha(f,n)$.

$A_\alpha(.5,n)$

α	n: 10	10^2	10^3	10^4	10^5	10^6
.1	[.196,.804]	[.393,.607]	[.466,.534]	[.489,.511]	[.496,.504]	[.498,.502]
.01	[.112,.888]	[.351,.649]	[.452,.548]	[.484,.516]	[.495,.505]	[.498,.502]
.001	[.068,.932]	[.320,.680]	[.441,.559]	[.481,.519]	[.494,.506]	[.498,.502]

$A_\alpha(.9,n)$

α	n: 10	10^2	10^3	10^4	10^5	10^6
.1	[.596,.996]	[.823,.953]	[.878,.919]	[.893,.907]	[.897,.903]	[.899,.901]
.01	[.446,1.00]	[.785,.967]	[.868,.927]	[.890,.909]	[.897,.903]	[.899,.901]
.001	[.338,1.00]	[.754,.976]	[.861,.932]	[.888,.911]	[.897,.903]	[.899,.901]

The statistical induction argument bears obvious similarities to orthodox statistical reasoning. It would be very surprising if that were not so, because all theories of inverse inference are based upon variants of the intuitive argument discussed above. But it should also be emphasized that if you try to list the exact points of similarity between the statistical induction argument and orthodox statistical reasoning concerning confidence intervals or significance testing, there are as many important differences as there are similarities. These include the use of indefinite probabilities in place of definite probabilities, the use of nonclassical direct inference in place of assumptions of independence, the *derivation* of the principle governing the use of likelihood ratios rather than the simple postulation of the likelihood principle, and the projectibility constraint. The latter deserves particular emphasis. A few writers have suggested general theories of statistical inference that are based on likelihood ratios and agree closely with the present account when applied to the special case of statistical induction. However, such theories have been based merely on statistical intuition, without an underlying rationale of the sort given here, and they are all subject to counterexamples having to do with projectibility.

Enumerative induction differs from statistical induction in that our inductive sample must consist of *B*'s *all of which are A*'s and our conclusion is the generalization "$B \Rightarrow$ A". It is shown in Pollock [1984a] and [1990] that in the special case in which the observed relative frequency is 1, our theory of nomic probability allows us to extend the statistical induction argument to obtain an argument for the conclusion "$B \Rightarrow$ A". Thus enumerative induction can also be justified on the basis of nomic probability.

6. Conclusions

To briefly recapitulate, the theory of nomic probability aims to make objective probability philosophically respectable by providing a mathematically precise theory of probabilistic reasoning. This theory has two primitive parts: (1) a set of computational principles comprising a strengthened probability calculus, and (2) an acceptance rule licensing defeasible inferences from high probabilities. Some of the computational principles are novel, but the resulting calculus of nomic probabilities is still just a strengthened probability calculus. The strengthenings are important, but not philosophically revolutionary. The main weight of the theory is borne by the acceptance rule. But it deserves to be emphasized that that rule is not revolutionary either. It is just a tidied-up version of the classical statistical syllogism. What makes the theory work where earlier theories did not is the explicit use of defeasible reasoning. It is the importation of state-of-the-art epistemology that infuses new life into the old ideas on which the theory of nomic probability is based.

The most remarkable aspect of the theory of nomic probability is that theories of direct inference and induction fall out as theorems. We need no new postulates in order to justify these essential kinds of probabilistic reasoning. In an important sense, this comprises a solution to the problem of induction. Of course, you cannot get something from nothing, so we are still making epistemological assumptions that Hume would have found distasteful, but in an important sense principles of induction have been derived from principles that are epistemologically more basic.

BIBLIOGRAPHY

Barnard, G. A.,

1949 Statistical inference. *J. Royal Statistical Society B*, II, 115-149.

1966 The use of the likelihood function in statistical practice. *Proceedings V Berkeley Symposium on Mathematical Statistics and Probability* I, 27-40.

1967 The Bayesian controversy in statistical inference. *J. Inst. Actuar.* 93, 229-269.

Birnbaum, Allan

1969 Concepts of statistical inference. In *Philosophy, Science, and Method*, ed. S. Morganbesser, *P*. Suppes, and M. White, 112-143. New York: St. Martin's Press.

Chisholm, Roderick

1957 *Perceiving*. Ithaca: Cornell University Press.

1977 *Theory of Knowledge*. Prentice Hall, second edition

1981 A version of foundationalism. *Midwest Studies in Philosophy* V (1981), 543-564.

Cox, D. *R*.

1958 Some problems connected with statistical inference. *Annals of Mathematical Statistics* 29, 357-363.

Dempster, A. P.

1964 On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association* 59, 56-66.

1966 Review of *The Logic of Statistical Inference*. *Journal of the American Statistical Association* 61, 1233-5.

Doyle, J.

1979 A truth maintainance system, *Artificial Intelligence* 12 (231-72).

Edwards, A. W. F.,

1972 *Likelihood*. Cambridge: Cambridge University Press.

Etherington, D. W.

1983 *Formalizing nonmonotonic reasoning systems*. Computer Science Tech. Report No. 83-1, University of British Columbia, Vancouver.

Fisher, *R*. A.,

1921 On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* I, part 4, 3-32.

1922 On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society* series A, vol. 222, 309-368.

Gillies, D. A.,

1973 *An Objective Theory of Probability*. London: Methuen.

Goodman, Nelson

1955 *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.

Hacking, Ian,

1965 *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Kolmogoroff, A. N.

1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin.

Kyburg, Henry, Jr.

1961 *Probability and the Logic of Rational Belief*. Middletown: Wesleyan University Press.

1970 Conjunctivitis. In *Induction, Acceptance, and Rational Belief*, edited by Marshall Swain. Dordrecht: Reidel.

1974 *The Logical Foundations of Statistical Inference*. Dordrecht: Reidel.

1983 The reference class. *Philosophy of Science* 50, 374-397.

McCarthy, J.,

1980 Circumscription — a form of nonmonotonic reasoning, *Artificial Intelligence* 13 (27-39).

1984 Applications of circumscription to formalizing common sense knowledge, *Proceedings of the nonmonotonic reasoning workshop*.

McDermott, D. V.

1982 Nonmonotonic logic II: monmonotonic modal theories, *J. ACM* 29 (1), 33-57.

McDermott, D. V. and J. Doyle

1980 Nonmonotonic logic I, *Artificial Intelligence* 13 (41-72).

Nute, D.

1986 *LDR: A logic for defeasible reasoning* (ACMC Research Report No. 01-0013). Athens: University of Georgia.

1987 Defeasible reasoning, *Proc. of the Twentieth Annual Hawaii Intl. Conf. on Systems Sciences.*

1988 Defeasible reasoning: a philosophical analysis in Prolog, in *Aspects of AI* (ed. J. Fetzer), Reidel, forthcoming.

Pollock, John

1967 Criteria and our Knowledge of the Material World, *The Philosophical Review*, 28-60.

1968 What is an Epistemological Problem, *American Philosophical Quarterly*, 183-190.

1971 Perceptual Knowledge, *Philosophical Review*, 287-319.

1974 *Knowledge and Justification*. Princeton: Princeton University Press.

1974a Subjunctive generalizations. *Synthese* 28, 199-214.

1976 *Subjunctive Reasoning*. Dordrecht: Reidel.

1983 A theory of direct inference. *Theory and Decision* 15, 29-96.

1983a Epistemology and probability. *Synthese* 55, 231-252.

1984 Foundations for direct inference. *Theory and Decision* 17, 221-256.

1984a A solution to the problem of induction. *Nous*,

1986 *Contemporary Theories of Knowledge*. Rowman and Littlefield, Totowa, N.J.

1986a The paradox of the preface. *Philosophy of Science* 53, 246-258.

1987 Defeasible reasoning. *Cognitive Science* 11, 481-518.

1987a Probability and proportions. *Theory and Decision: Essays in Honor of Werner Leinfellner*, ed. H. Berghel and G. Eberlein. Reidel, Amsterdam.

1990 *Nomic Probability and the Foundations of Induction*, Oxford University Press.

1995 *Cognitive Carpentry*, MIT Press.

Poole, D.

1985 On the comparison of theories: preferring the most specific explanation, *Proc. of the Intl. Joint Conf. on AI*, Los Angeles.

Pratt, John W.

1961 Review of E. L. Lehmann's *Testing Statistical Hypotheses*. *Journal of the American Statistical Association* 56, 166.

Reiter, R.

1980 A logic for default reasoning, *Artificial Intelligence* 13 (81-132).

Seidenfeld, Teddy

1979 *Philosophical Problems of Statistical Inference*. Dordrecht: Reidel.