

# Direct Inference and Probable Probabilities<sup>1</sup>

John L. Pollock  
Department of Philosophy  
University of Arizona  
Tucson, Arizona 85721  
[pollock@arizona.edu](mailto:pollock@arizona.edu)  
<http://www.u.arizona.edu/~pollock>

## Abstract

New results in the theory of nomic probability have led to a theory of probable probabilities, which licenses defeasible inferences between probabilities that are not validated by the probability calculus. Among these are classical principles of direct inference together with some new more general principles that greatly strengthen direct inference and make it much more useful.

I first learned about direct inference from Henry Kyburg, when I was a young faculty member at the University of Rochester. I worked through his book *The Foundations of Statistical Inference* (Kyburg 1974) with great care, and had many long and fruitful conversations with him. Over the years, my own views evolved and diverged from Kyburg's, but his original insights are evident throughout my work.

Direct inference pertains to a fundamental but often overlooked distinction between two kinds of probabilities. What I will call *generic probabilities*<sup>2</sup> are general probabilities, relating properties or relations. For example, we can talk about the probability of an adult male of Slavic descent being lactose intolerant. This is not about any particular person — it expresses a relationship between the property of being an adult male of Slavic descent and the property of being lactose intolerant. Most forms of statistical inference or statistical induction are most naturally viewed as giving us information about generic probabilities. On the other hand, for many purposes we are more interested in propositions that are about particular persons, or more generally, about specific matters of fact. For example, in deciding how to treat Herman, an adult male of Slavic descent, his doctor may want to know the probability that Herman is lactose intolerant. This illustrates the need for a kind of probability that attaches to propositions rather than relating properties and relations. These are sometimes called “single case probabilities”, although that terminology is not very good because such probabilities can attach to propositions of any logical form. For example, we can ask how probable it is that there are no human beings over the age of 130. In the past, I called these “definite probabilities”, but now I will refer to them as *singular probabilities*.

Most people find the distinction between generic probabilities and singular probabilities intuitively clear, and yet most contemporary work in probability theory ignores it. This may be because of the popularity of subjective probability, which does not seem to be able to make sense of generic probabilities.<sup>3</sup> This neglect of generic probabilities is odd, however, because most of the historical work on objective probability pertained to generic probabilities. I will say more about this below.

If statistical inference leads to knowledge of generic probabilities, we still need a way of discovering the values of singular probabilities. Direct inference is supposed to fill this gap. For

---

<sup>1</sup> This work was supported by NSF grant no. IIS-0412791.

<sup>2</sup> In the past, I followed Jackson and Pargetter 1973 in calling these “indefinite probabilities”, but I never liked that terminology.

<sup>3</sup> The only attempt I know to define generic probability within a subjectivist framework is that of Jackson and Pargetter (1973), but it has the consequence that if the agent knows the relative frequency then the generic probability equals the relative frequency.

instance, if we know that the generic probability of an adult male of Slavic descent being lactose intolerant is .35, and we know that Herman is an adult male of Slavic descent, we may infer that the probability of Herman being lactose intolerant is .35. This illustrates direct inference. In direct inference we infer singular probabilities from generic probabilities.

Kyburg (1974) was the first to propose a detailed theory of direct inference. His theory proceeded by postulating precise rules of direct inference and then defining singular probability in terms of those rules. In Pollock (1990), I proposed instead to define singular probabilities in terms of generic probabilities and then derive rules of direct inference from the theory of generic probability. The rules thus obtained were similar to Kyburg's, but like Kyburg's theory, the theory retained a somewhat ad hoc flavor. Some of the assumptions made about generic probabilities were chosen specifically to obtain the desired rules of direct inference. This paper aims to rectify that. New results in the theory of generic probability lead to a general theory of "probable probabilities", and that leads to a more general and better motivated theory of direct inference.

## 1. Generic Probabilities and Nomic Probability

Most objective approaches tie probabilities to relative frequencies in some essential way, and the resulting probabilities have the same logical form as the relative frequencies. That is, they are generic probabilities. The simplest theories identify generic probabilities with relative frequencies.<sup>4</sup> This was Kyburg's (1961, 1974) view. However, it is often objected, fairly I think, that such "finite frequency theories" are at least sometimes inadequate because our probability judgments often diverge from relative frequencies. For example, we can talk about a coin being fair (and so the generic probability of a flip landing heads is 0.5) even when it is flipped only once and then destroyed (in which case the relative frequency is either 1 or 0). For understanding such generic probabilities, it has been suggested that we need a notion of probability that talks about *possible* instances of properties as well as actual instances. Theories of this sort are sometimes called "hypothetical frequency theories". C. S. Peirce was perhaps the first to make a suggestion of this sort. Similarly, the statistician R. A. Fisher, regarded by many as "the father of modern statistics", identified probabilities with ratios in a "hypothetical infinite population, of which the actual data is regarded as constituting a random sample" (1922, p. 311). Karl Popper (1956, 1957, and 1959) endorsed a theory along these lines and called the resulting probabilities *propensities*. Henry Kyburg (1974a) was the first to construct a precise version of this theory (although he did not endorse the theory), and it is to him that we owe the name "hypothetical frequency theories". Kyburg (1974a) also insisted that von Mises should be considered a hypothetical frequentist. More recent attempts to formulate precise versions of what might be regarded as hypothetical frequency theories are van Fraassen (1981), Bacchus (1990), Halpern (1990), Pollock (1983, 1984, 1990), and Bacchus et al (1996). I will sketch my own proposal here.

I do not think that it should be supposed that there is just one sensible kind of generic probability. However, in my (1990) I suggested that there is a central kind of generic probability in terms of which a number of other kinds can be defined. This central kind of generic probability is what I called *nomic probability*. Nomic probabilities are supposed to be the subject matter of statistical laws of nature. Exceptionless general laws, like "All electrons are negatively charged", are not just about actual electrons, but also about all physically possible electrons. We can think of such a law as reporting that any physically possible electron would be negatively charged. This is an example of a *nomic generalization*. We can think of nomic probabilities as telling us instead that that a certain proportion of physically possible objects of a certain sort will have some other property. For example, we might have a law to the effect that the probability of a hadron being negatively charged is .5. We can think of this as telling us that half of all physically possible hadrons would be negatively charged.

Following Pollock (1990), I propose that we can identify the *nomic probability*  $\text{prob}_x(Fx/Gx)$  with the proportion of physically possible  $G$ 's that are  $F$ 's. A *physically possible*  $G$  is defined to be an ordered pair  $\langle w, x \rangle$  such that  $w$  is a physically possible world (one compatible with all of the physical laws) and  $x$  has the property  $G$  at  $w$ . Let us define the *subproperty relation* as follows:

$F \leq G$  iff it is physically necessary (follows from true physical laws) that  $(\forall x)(Fx \rightarrow Gx)$ .

<sup>4</sup> Examples are Russell (1948); Braithwaite (1953); Kyburg (1961, 1974); Sklar (1970, 1973). William Kneale (1949) traces the frequency theory to R. L. Ellis, writing in the 1840's, and John Venn (1888) and C. S. Peirce in the 1880's and 1890's.

We can think of the subproperty relation as a kind of nomic entailment relation (holding between properties rather than propositions). More generally,  $F$  and  $G$  can have any number of free variables (not necessarily the same number), in which case  $F \leq G$  iff the universal closure of  $(F \rightarrow G)$  is physically necessary.

Given a suitable proportion function  $\rho$ , we could stipulate that, where  $\mathfrak{F}$  and  $\mathfrak{G}$  are the sets of physically possible  $F$ 's and  $G$ 's respectively:

$$\text{prob}_x(Fx/Gx) = \rho(\mathfrak{F}, \mathfrak{G}).^5$$

However, it is unlikely that we can pick out the right proportion function without appealing to prob itself, so the postulate is simply that *there is* some proportion function related to prob as above. This is merely taken to tell us something about the formal properties of prob. Rather than axiomatizing prob directly, it turns out to be more convenient to adopt axioms for the proportion function. Proportion functions are a generalization of measure functions, studied in mathematics in measure theory. Pollock (1990) showed that, given the assumptions adopted there,  $\rho$  and prob are interdefinable, so the same empirical considerations that enable us to evaluate prob inductively also determine  $\rho$ .

Note that  $\text{prob}_x$  is a variable-binding operator, binding the variable  $x$ . When there is no danger of confusion, I will omit the subscript " $x$ ", but sometimes we will want to quantify into probability contexts, in which case it will be important to distinguish between the variables bound by "prob" and those that are left free. To simplify expressions, I will often omit the variables, writing " $\text{prob}(F/G)$ " for " $\text{prob}(Fx/Gx)$ " when no confusion will result.

It is often convenient to write proportions in the same logical form as probabilities, so where  $\varphi$  and  $\theta$  are open formulas with free variable  $x$ , let  $\rho_x(\varphi/\theta) = \rho(\{x \mid \varphi \& \theta\}, \{x \mid \theta\})$ .  $\rho_x$  is a variable-binding operator, binding the variable  $x$ . Again, when there is no danger of confusion, I will typically omit the subscript " $x$ ".

I will make three classes of assumptions about the proportion function. Let  $\#X$  be the cardinality of a set  $X$ . If  $Y$  is finite, I assume:

$$\rho(X, Y) = \frac{\#X \cap Y}{\#Y}.$$

However, for present purposes the proportion function is most useful in talking about proportions among infinite sets. The sets  $\mathfrak{F}$  and  $\mathfrak{G}$  will invariably be infinite, if for no other reason than that there are infinitely many physically possible worlds in which there are  $F$ 's and  $G$ 's.

My second set of assumptions is that standard axioms for conditional probabilities hold for proportions. These axioms automatically hold for relative frequencies among finite sets, so the assumption is just that they also hold for proportions among infinite sets.

Because proportions are always conditional, they cannot be defined in the standard way in terms of ratios of "unconditional proportions". The latter notion makes no sense. Technically, proportions are *Popper functions* (Popper 1956). This has two important consequences. First,  $\rho(A/B \& C)$  can be well-defined even when  $\rho(B/C) = 0$ . Second, if  $\rho(A/B) = 1$ , it does not follow that  $\rho(A/B \& C) = 1$ . That can fail when  $\rho(C/B) = 0$ . Thus, for example,

$$\text{prob}(2x \text{ is irrational} / x \text{ is a real number}) = 1$$

but

$$\text{prob}(2x \text{ is irrational} / x \text{ is a real number} \& x \text{ is rational}) = 0.$$

That further assumptions are needed derives from the fact that the standard probability calculus is a calculus of singular probabilities rather than generic probabilities. A calculus of generic probabilities is related to the calculus of singular probabilities in a manner roughly analogous to the relationship between the predicate calculus and the singular calculus. Thus we get some principles pertaining specifically to relations that hold for generic probabilities but cannot even be formulated

<sup>5</sup> Probabilities relating  $n$ -place relations are treated similarly. I will generally just write the one-variable versions of various principles, but they generalize to  $n$ -variable versions in the obvious way.

in the standard probability calculus. For instance, Pollock (1990) endorsed the following two principles:

**Individuals:**

$$\text{prob}(Fxy/Gxy \ \& \ y = a) = \text{prob}(Fxa/Gxa)$$

**PPROB:**

$$\text{prob}(Fx/Gx \ \& \ \text{prob}(Fx/Gx) = r) = r.$$

I will not use or assume either of these principles in this paper, but I mention them just to illustrate that there are reasonable-seeming principles governing generic probabilities that are not even well formed in the standard probability calculus.

What I do need in the present paper is three assumptions about proportions that go beyond merely imposing the standard axioms for the probability calculus. The three assumptions I will make are:

**Finite Set Principle:**

For any set  $B$ ,  $N > 0$ , and open formula  $\Phi$ ,

$$\rho_X(\Phi(X) / X \subseteq B \ \& \ \#X = N) =$$

$$\rho_{x_1, \dots, x_N}(\Phi(\{x_1, \dots, x_N\}) / x_1, \dots, x_N \text{ are pairwise distinct} \ \& \ x_1, \dots, x_N \in B).$$

**Projection Principle:**

If  $0 \leq p, q \leq 1$  and  $(\forall y)(Gy \rightarrow \rho_x(Fx/Rxy) \in [p, q])$ , then  $\rho_{x,y}(Fx/Rxy \ \& \ Gy) \in [p, q]$ .<sup>6</sup>

**Crossproduct Principle:**

If  $C$  and  $D$  are nonempty,  $\rho(A \times B, C \times D) = \rho(A, C) \cdot \rho(B, D)$ .

Note that these three principles are all theorems of elementary set theory when the sets in question are finite. For instance, the crossproduct principle holds for finite sets because  $\#(A \times B) = (\#A) \cdot (\#B)$ , and hence

$$\begin{aligned} \rho(A \times B, C \times D) &= \frac{\#((A \times B) \cap (C \times D))}{\#(C \times D)} = \frac{\#((A \cap C) \times (B \cap D))}{\#(C \times D)} \\ &= \frac{\#(A \cap C) \cdot \#(B \cap D)}{\#C \cdot \#D} = \frac{\#(A \cap C)}{\#C} \cdot \frac{\#(B \cap D)}{\#D} = \rho(A, C) \cdot \rho(B, D). \end{aligned}$$

My assumption is simply that  $\rho$  continues to have these algebraic properties even when applied to infinite sets. I take it that this is a fairly conservative set of assumptions.

I often hear the objection that in affirming the Crossproduct Principle, I must be making a hidden assumption of statistical independence. However, that is to confuse proportions with probabilities. The Crossproduct Principle is about proportions — not probabilities. For finite sets, proportions are computed by simply counting members and computing ratios of cardinalities. It makes no sense to talk about statistical independence in this context. For infinite sets we cannot just count members any more, but the algebra is the same. It is because the algebra of proportions is simpler than the algebra of probabilities that it is useful to axiomatize nomic probabilities indirectly by adopting axioms for proportions.

## 2. The Statistical Syllogism

Pollock (1990) derived the entire epistemological theory of nomic probability from a single epistemological principle coupled with a mathematical theory that amounts to a calculus of nomic probabilities. The single epistemological principle that underlies probabilistic reasoning is the *statistical syllogism*. Although orthodox Bayesians would have us think otherwise, there are some

---

<sup>6</sup> Note that this is a different (and more conservative) principle than the one called “Projection” in Pollock (1990).

things that we simply believe, without attaching a probability to them. We often form beliefs on the basis of high generic probabilities. For example, I believe that today is March 31 because that is the date displayed on my watch. I do not believe that my watch is always right. At best, the generic probability is high of the date being as displayed on my watch, but that seems to be enough to justify me, at least provisionally (or defeasibly) in believing that this is March 31. This illustrates the use of *the statistical syllogism*. The statistical syllogism licenses inferences from high generic probabilities and singular propositions. Crudely, construing “most” as expressing a generic probability, the statistical syllogism has the form “Most A’s are B’s, and this is an A, so (defeasibly) this is a B.” I will use the following precise form of the statistical syllogism:

### **Statistical Syllogism:**

If  $F$  is projectible with respect to  $G$  and  $r > 0.5$ , then  $\lceil Gc \ \& \ \text{prob}(F/G) \geq r \rceil$  is a defeasible reason for  $\lceil Fc \rceil$ , the strength of the reason being a monotonic increasing function of  $r$ .

Throughout, I assume the theory of defeasible reasoning described in Pollock (1995, 2008a). However, most of the details will not matter for present purposes.

I take it that the statistical syllogism is a very intuitive principle, and it is clear that we employ it constantly in our everyday reasoning. Suppose, for example, that you read in the newspaper that George Bush is visiting Guatemala, and you believe what you read. What justifies your belief? No one believes that everything printed in the newspaper is true. What you believe is that certain kinds of reports published in certain kinds of newspapers tend to be true, and this report is of that kind. It is the statistical syllogism that justifies your belief.

The projectibility constraint in the statistical syllogism is the familiar projectibility constraint on inductive reasoning, first noted by Goodman (1955). Kyburg (1974) was the first to observe the need for some constraint on what properties the statistical syllogism can be applied to, although he did not initially identify the constraint with the standard projectibility constraint. One might wonder what projectibility has to do with the statistical syllogism. But it was argued in (Pollock 1990), on the strength of what were taken to be intuitively compelling examples, that the statistical syllogism must be so constrained. Furthermore, it was shown that without a projectibility constraint, the statistical syllogism is self-defeating, because for any intuitively correct application of the statistical syllogism it is possible to construct a conflicting (but unintuitive) application leading to a contrary conclusion. This is the same problem that Goodman first noted in connection with induction. Pollock (1990) then went on to argue that the reason the constraint on the statistical syllogism is the same as the one on induction is that principles of induction can be derived from the calculus of nomic probabilities with the aid of the statistical syllogism, so the projectibility constraint on induction derives from that on the statistical syllogism. The projectibility constraint is important, but also problematic because no one has a good analysis of it. I will not discuss it further here, but I will assume that it is satisfied in the places in which I use the statistical syllogism.

The statistical syllogism is a defeasible inference scheme, so it is subject to defeat. I now believe that the only primitive (underived) principle of defeat required for the statistical syllogism is that of subproperty defeat:

### **Subproperty Defeat for the Statistical Syllogism:**

If  $H$  is projectible with respect to  $G$ , then  $\lceil Hc \ \& \ \text{prob}(F/G\&H) < \text{prob}(F/G) \rceil$  is an undercutting defeater for the inference by the statistical syllogism from  $\lceil Gc \ \& \ \text{prob}(F/G) \geq r \rceil$  to  $\lceil Fc \rceil$ .<sup>7</sup>

In other words, information about  $c$  that lowers the probability of its being  $F$  constitutes a defeater. Note that if  $\text{prob}(F/G\&H)$  is less than  $r$  but still high, one may still be able to make a weaker inference to the conclusion that  $Fc$ , but from the distinct premise  $\lceil Gc \ \& \ \text{prob}(F/G\&H) = s \rceil$ .

Pollock (1990) argued that we need additional defeaters for the statistical syllogism besides subproperty defeaters, and formulated several candidates for such defeaters. But one of the conclusions of the research described in this paper and in Pollock (2008) is that the additional

---

<sup>7</sup> There are two kinds of defeaters. Rebutting defeaters attack the conclusion of an inference, and undercutting defeaters attack the inference itself without attacking the conclusion. Here I assume some form of the OSCAR theory of defeasible reasoning (Pollock 1995). For a sketch of that theory see Pollock (2008a).

defeaters can all be viewed as derived defeaters, with subproperty defeaters being the only primitive defeaters for the statistical syllogism.

### 3. Singular Probabilities

One of the central theses of the theory of nomic probability is that principles of direct inference can be derived as theorems from the calculus of nomic probabilities together with the statistical syllogism. This is a notable divergence from Kyburg (1974), who took the principles of direct inference to be primitive and used them to define singular probabilities. My approach is instead to define singular probabilities directly, in terms of nomic probabilities, and then prove a collection of mathematical theorems that show how to defeasibly infer the values of singular probabilities from the values of related nomic probabilities.

There is more than one kind of singular probability. “Objective chance” is a purely objective singular probability pertaining, for example, to chance events in quantum mechanics. In Pollock (1990) it was suggested that there is a theory of direct inference applicable to these probabilities, however in this paper I will focus instead on the application of direct inference to “mixed physical-epistemic probabilities”. The probabilities used in practical deliberation must have a strong epistemic element. For example, if I am deciding whether to carry an umbrella when I go to work today, before looking outside I may know that the generic probability of rain in Tucson this time of year is only .05, and so I will conclude that it is extremely unlikely to rain today. But then I look out the window and see huge dark clouds looming overhead. At that point I conclude that rain is likely. This example illustrates that the probability I employ in deciding whether to carry an umbrella is partly a function of nomic probabilities, but it changes as my knowledge of the situation changes. I propose to capture this by identifying singular probabilities with a special class of generic probabilities. The present treatment is a generalization of that given in my (1984 and 1990).<sup>8</sup> Let  $\mathbf{K}$  be the conjunction of all the propositions the agent knows to be true, and let  $\mathfrak{R}$  be the set of all physically possible worlds at which  $\mathbf{K}$  is true (“ $\mathbf{K}$ -worlds”). I propose that we define the singular probability  $\text{PROB}(P)$  to be the proportion of  $\mathbf{K}$ -worlds at which  $P$  is true. Where  $\mathfrak{P}$  is the set of all  $P$ -worlds:

$$\text{PROB}(P) = \rho(\mathfrak{P}, \mathfrak{R}).$$

More generally, where  $\mathfrak{Q}$  is the set of all  $Q$ -worlds, we can define:

$$\text{If } \mathfrak{Q} \cap \mathfrak{R} \neq \emptyset \text{ then } \text{PROB}(P/Q) = \rho(\mathfrak{P}, \mathfrak{Q} \cap \mathfrak{R}).$$

Formally, this is analogous to Carnap’s (1950,1952) logical probability, with the important difference that Carnap took  $\rho$  to be specified logically, whereas I take the identity of  $\rho$  to be a contingent fact.  $\rho$  is determined by the values of contingently true nomic probabilities, and their values are discovered by various kinds of statistical induction.

It turns out (and is proven in Pollock 2008) that singular probabilities, so defined, can be identified with a special class of nomic probabilities:

#### Representation Theorem for Singular Probabilities:

- (1)  $\text{PROB}(Fa) = \text{prob}(Fx/x = a \ \& \ \mathbf{K})$ ;
- (2) If it is physically necessary that  $[\mathbf{K} \rightarrow (Q \leftrightarrow Sa_1 \dots a_n)]$  and that  $[(Q \& \mathbf{K}) \rightarrow (P \leftrightarrow Ra_1 \dots a_n)]$ , and  $Q$  is consistent with  $\mathbf{K}$ , then  $\text{PROB}(P/Q) = \text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \ \& \ x_1 = a_1 \ \& \ \dots \ \& \ x_n = a_n \ \& \ \mathbf{K})$ .
- (3)  $\text{PROB}(P) = \text{prob}(P \ \& \ x=x/x = x \ \& \ \mathbf{K})$ .

$\text{PROB}(P)$  is a kind of “mixed physical/epistemic probability”, because it combines background knowledge in the form of  $\mathbf{K}$  with generic probabilities.<sup>9</sup>

The probability  $\text{prob}(Fx/x = a \ \& \ \mathbf{K})$  is a peculiar-looking nomic probability. It is an generic probability, because “ $x$ ” is a free variable, but the probability is only about one object. As such it

<sup>8</sup> Bacchus (1990) gave a somewhat similar account of direct inference, drawing on my 1983 and 1984.

<sup>9</sup> See chapter six of Pollock (2006) for further discussion of these mixed physical/epistemic probabilities.

cannot be evaluated by statistical induction or other familiar forms of statistical reasoning. However, as I will argue next, it can be evaluated using direct inference.

## 4. Probable Probabilities

In Pollock (1990), I derived principles of direct inference from some very complex (and sometimes dubious) postulates about the proportion function  $\rho$ . However, I have recently found a much better way of proceeding. If we make just the simple and presumably uncontroversial assumptions about  $\rho$  that were enumerated in section one, we can prove some very general theorems about nomic probabilities that are of far-reaching importance. These are theorems about “probable probabilities”. If we know the probabilities of a few simple propositions, the probability calculus is generally too weak to allow us to compute more than rather broad bounds on the probabilities of logical compounds of those propositions. For example, suppose we know that  $\text{PROB}(P) = \text{PROB}(R) = .7$  and  $\text{PROB}(Q) = \text{PROB}(S) = .6$ . What can we conclude about  $\text{PROB}(P \& Q)$ ? All the probability calculus enables us to infer is that  $.3 \leq \text{PROB}(P \& Q) \leq .6$ . That does not tell us much. Similarly, all we can conclude about  $\text{PROB}(P \vee Q)$  is that  $.7 \leq \text{PROB}(P \vee Q) \leq 1.0$ . And all we can conclude about  $((P \& Q) \vee (R \& S))$  is that  $.3 \leq \text{PROB}((P \& Q) \vee (R \& S)) \leq 1.0$ . In general, the probability calculus imposes fairly weak constraints on the probabilities of logical compounds, but it falls far short of enabling us to compute unique values. But the theory of probable probabilities shows that, under very general circumstances, there are precisely computable values that we can defeasibly expect these unknown probabilities to have. The unknown probabilities are not logically guaranteed to have those values, but the sets of cases in which they do not are of measure 0. Among the consequences of these theorems about probable probabilities will be some principles of direct inference.

The fundamental theorems about probable probabilities are proven in Pollock (2008), so I will just state them here. Given a list of variables  $X_1, \dots, X_n$  ranging over subsets of a set  $U$ , Boolean compounds of these sets are compounds formed by union, intersection, and set-complement. So, for example  $(X \cup Y) - Z$  is a Boolean compound of  $X$ ,  $Y$ , and  $Z$ . *Linear constraints* on the Boolean compounds either state the values of certain proportions, e.g., stipulating that  $\rho(X, Y) = r$ , or they relate proportions using linear equations. For example, if we know that  $X = Y \cup Z$ , that generates the linear constraint

$$\rho(X, U) = \rho(Y, U) + \rho(Z, U) - \rho(X \cap Z, U).$$

Where “ $x \approx_{\delta} y$ ” means that the absolute value of  $x - y$  is less than  $\delta$ , my central theorem is the following purely combinatorial theorem about finite sets:

### **Probable Proportions Theorem:**

Let  $U, X_1, \dots, X_n$  be a set of variables ranging over sets, and consider a finite set  $LC$  of linear constraints on proportions between Boolean compounds of those variables. If  $LC$  is consistent with the probability calculus, then for any pair of Boolean compounds  $P, Q$  of  $U, X_1, \dots, X_n$  there is a real number  $r$  between 0 and 1 such that for every  $\varepsilon, \delta > 0$ , there is an  $N$  such that if  $U$  is finite and  $\#U > N$ , then

$$\rho_{X_1, \dots, X_n} \left( \rho(P, Q) \approx_{\delta} r \mid LC \ \& \ X_1, \dots, X_n \subseteq U \right) \geq 1 - \varepsilon.$$

This is a straightforward theorem of set theory, and it is proven in Pollock (2008). It is to be emphasized that this is just a mathematical theorem about finite sets, and does not depend upon any of our assumptions about  $\rho$  except for the assumption that for finite sets,  $\rho(X, Y) = \frac{\#X \cap Y}{\#Y}$ .

This combinatorial theorem underlies all of the principles developed in this paper.

Given the rather simple assumptions I made about  $\rho$  in section one, we can use familiar-looking mathematics to prove (Pollock 2008):

### **Law of Large Numbers for Proportions:**

If  $B$  is infinite and  $\rho(A/B) = p$  then for every  $\varepsilon, \delta > 0$ , there is an  $N$  such that

$$\rho_X \left( \rho(A/X) \approx_{\delta} p \mid X \subseteq B \ \& \ X \text{ is finite} \ \& \ \#X \geq N \right) \geq 1 - \varepsilon.$$

Note that unlike Laws of Large Numbers for probabilities, the Law of Large Numbers for Proportions does not require an assumption of statistical independence. This is because it is derived from the crossproduct principle, and as remarked in section one, no such assumption is required (or even intelligible) for the crossproduct principle.

Given the law of large numbers for proportions, we can prove:

### **Limit Principle for Proportions:**

Consider a finite set  $LC$  of linear constraints on proportions between Boolean compounds of a list of variables  $U, X_1, \dots, X_n$ . For any real number  $r$  between 0 and 1, if for every  $\varepsilon, \delta > 0$ , if there is an  $N$  such that for any finite set  $U$  such that  $\#U > N$ ,

$$\rho_{X_1, \dots, X_n} \left( \rho(P, Q) \approx_{\delta} r \mid LC \ \& \ X_1, \dots, X_n \subseteq U \right) \geq 1 - \varepsilon,$$

then for any infinite set  $U$ , for every  $\delta > 0$ :

$$\rho_{X_1, \dots, X_n} \left( \rho(P, Q) \approx_{\delta} r \mid LC \ \& \ X_1, \dots, X_n \subseteq U \right) = 1.$$

Nomic probabilities are proportions among physically possible objects. For any property  $F$  that is not extraordinarily contrived, the set  $\mathfrak{F}$  of physically possible  $F$ 's will be infinite.<sup>10</sup> Thus, given the limit principle for proportions, the Probable Proportions Theorem entails:

### **Probable Probabilities Principle:**

Let  $U, X_1, \dots, X_n$  be a set of variables ranging over properties and relations, and consider a finite set  $LC$  of linear constraints on probabilities between truth-functional compounds of those variables. If  $LC$  is consistent with the probability calculus, then for any pair of truth-functional compounds  $P, Q$  of  $U, X_1, \dots, X_n$  there is a real number  $r$  between 0 and 1 such that for every  $\delta > 0$ ,

$$\text{prob}_{X_1, \dots, X_n} \left( \text{prob}(P/Q) \approx_{\delta} r \mid LC \ \& \ X_1, \dots, X_n \subseteq U \right) = 1.$$

Next note that we can apply the statistical syllogism to the probability formulated in the expectable probabilities principle (I assume without discussion that the projectibility constraint is satisfied). For every  $\delta > 0$ , this gives us a defeasible reason for expecting that if  $LC$  then  $\text{prob}(P/Q) \approx_{\delta} r$ , and these conclusions jointly entail that  $\text{prob}(P/Q) = r$ . Thus we are led to a defeasible inference scheme:

### **Expectable Probabilities Principle:**

Let  $U, X_1, \dots, X_n$  be a set of variables ranging over sets, and consider a finite set  $LC$  of linear constraints on proportions between Boolean compounds of those variables. Then for any pair of Boolean compounds  $P, Q$  of  $U, X_1, \dots, X_n$  if  $r$  is a real number between 0 and 1 such that for every  $\varepsilon, \delta > 0$ , there is an  $N$  such that if  $U$  is finite and  $\#U > N$ , then

$$\rho_{X_1, \dots, X_n} \left( \rho(P, Q) \approx_{\delta} r \mid LC \ \& \ X_1, \dots, X_n \subseteq U \right) \geq 1 - \varepsilon,$$

then for any properties  $X_1, \dots, X_n$  it is defeasibly reasonable to expect that  $\text{prob}(P/Q) = r$ .

I will refer to  $r$  as *the expectable value of  $\text{prob}(P/Q)$* . The Expectable Probabilities Principle establishes the existence of expectable values for probabilities under very general circumstances.

<sup>10</sup> The following principles apply only to properties for which there are infinitely many physically possible instances, but I will not explicitly include the qualification "non-contrived" in the principles.



The preceding principles tell us that expectable values exist. It turns out that there is a general strategy for finding and proving theorems describing these expectable values, and I have written a computer program (in Common LISP) that will often do this automatically, both finding the theorems and producing human readable proofs. It can be downloaded from [http://oscarhome.soc-sci.arizona.edu/ftp/OSCAR-web-page/CODE/Code for probable probabilities.zip](http://oscarhome.soc-sci.arizona.edu/ftp/OSCAR-web-page/CODE/Code%20for%20probable%20probabilities.zip). I will refer to this as *the probable probabilities software*.

To illustrate the power of the Expectable Probabilities Principle, I list a few instances of it. The proofs can be found in Pollock (2008), or obtained directly by running the aforementioned program. First, consider statistical independence. Probability practitioners almost invariably assume (at least defeasibly) that probabilities are statistically independent unless they have some concrete reason for thinking otherwise. However, they often do so apologetically, because although such an assumption seems intuitively right, its justification has heretofore eluded the probability community. However, the theory of probable probabilities puts a defeasible assumption of statistical independence on a firm mathematical footing. Our first theorem about expectable values is:

**Principle of Expectable Statistical Independence:**

If  $\text{prob}(A/C) = r$  and  $\text{prob}(B/C) = s$ , the expectable value of  $\text{prob}(A \& B/C) = r \cdot s$ .

This, I take it, explains the commonly held intuition that we should be able to assume statistical independence when we see no reason for thinking it fails. This is a purely mathematical justification of the defeasible assumption.

Perhaps more important for practical purposes, the theory of probable probabilities gives us the mathematical wherewithal to investigate the circumstances under which we should not assume statistical independence. These amount to defeaters for the principle of expectable statistical independence. The precise forms of these defeaters are not initially intuitively obvious, and some of them are rather surprising. Let us begin with a defeater that does have some claim to being intuitive, and turn to the less intuitive ones as we proceed. We can prove the following generalization of the principle of statistical independence:

**Principle of Statistical Independence with Overlap:**

If  $\text{prob}(A/C) = r$ ,  $\text{prob}(B/C) = s$ ,  $\text{prob}(D/C) = g$ ,  $(A \& C) \leq D$ , and  $(B \& C) \leq D$ , then the expectable

value of  $\text{prob}(A \& B/C) = \frac{r \cdot s}{g}$ .

To illustrate with a simple and intuitive case, suppose  $A = A_0 \& D$  and  $B = B_0 \& D$ . Given no reason to think otherwise, we would expect  $A_0$ ,  $B_0$ , and  $D$  to be statistically independent. But then we would expect that

$$\begin{aligned} \text{prob}(A \& B/C) &= \text{prob}(A_0 \& D \& B_0/C) = \text{prob}(A_0/C) \cdot \text{prob}(D/C) \cdot \text{prob}(B_0/C) \\ &= \frac{\text{prob}(A_0 \& D/C) \cdot \text{prob}(B_0 \& D/C)}{\text{prob}(D/C)} = \frac{r \cdot s}{g}. \end{aligned}$$

The principle of statistical independence with overlap generates a defeater for the Principle of Expectable Statistical Independence. The instance of the Probable Proportions Theorem that underlies the Principle of Statistical Independence is the following:

$$\text{prob}_{X,Y,Z} \left( \begin{array}{l} \text{prob}(X \& Y/Z) \approx_{\delta} r \cdot s / \\ X, Y, Z \leq U \text{ and } \text{prob}(X/Z) = r \text{ and } \text{prob}(Y/Z) = s \\ \text{and } \text{prob}(X/U) = \alpha \text{ and } \text{prob}(Y/U) = \beta \text{ and } \text{prob}(Z/U) = \gamma \end{array} \right) = 1.$$

On the other hand, the instance of the Probable Proportions Theorem that underlies the Principle of Statistical Independence with Overlap is:

$$\text{prob}_{X,Y,Z,W} \left( \begin{array}{l} \text{prob}(X \& Y / Z) \approx \frac{r \cdot s}{\delta} / \\ X, Y, Z, W \leq U \text{ and } \text{prob}(X / Z) = r \text{ and } \text{prob}(Y / Z) = s \\ \text{and } (X \& Z) \leq W \text{ and } (Y \& Z) \leq \text{ and } \text{prob}(W / Z) = g \text{ and } \text{prob}(W / U) = \zeta \\ \text{and } \text{prob}(X / U) = \alpha \text{ and } \text{prob}(Y / U) = \beta \text{ and } \text{prob}(Z / U) = \gamma \end{array} \right) = 1.$$

The latter probability takes account of more information than the former, so it provides a subproperty defeater for the use of the statistical syllogism:

**Overlap Defeat for Statistical Independence:**

“ $(A \& C) \leq D$ ,  $(B \& C) \leq D$ , and  $\text{prob}(D/C) \neq 1$ ” is an undercutting defeater for the inference from “ $\text{prob}(A/C) = r$  and  $\text{prob}(B/C) = s$ ” to “ $\text{prob}(A \& B/C) = r \cdot s$ ” by the Principle of Statistical Independence.

The principle of overlap defeat can seem surprising. Suppose you know that  $\text{prob}(A/C) = r$  and  $\text{prob}(B/C) = s$ , and are inclined to infer that  $\text{prob}(A \& B/C) = r \cdot s$ . As long as  $r, s < 1$ , there will always be a  $D$  such that  $(A \& C) \leq D$ ,  $(B \& C) \leq D$ , and  $\text{prob}(D/C) \neq 1$ . Does this mean that the inference is always defeated? It does not, but understanding why is a bit complicated. First, what we know in general is the existential generalization  $(\exists D)[(A \& C) \leq D \text{ and } (B \& C) \leq D \text{ and } \text{prob}(D/C) \neq 1]$ . But the defeater requires knowing of a specific such  $D$ . The reason for this is that it is not true in general that  $\text{prob}(Fx/Rxy) = \text{prob}(Fx/(\exists y)Rxy)$ . For example, let  $Fx$  be “ $x = 1$ ” and let  $Rxy$  be “ $x < y \& x, y$  are natural numbers  $\leq 2$ ”. Then  $\text{prob}(Fx/Rxy) = 1/3$ , but  $\text{prob}(Fx/(\exists y)Rxy) = 1/2$ . Accordingly, we cannot assume that

$$\begin{aligned} & \text{prob}_{X,Y,Z,W} \left( \begin{array}{l} \text{prob}(X \& Y / Z) \approx \frac{r \cdot s}{\delta} / \\ X, Y, Z, W \leq U \text{ and } \text{prob}(X / Z) = r \text{ and } \text{prob}(Y / Z) = s \text{ and } \\ (X \& Z) \leq W \text{ and } (Y \& Z) \leq \text{ and } \text{prob}(W / Z) = g \text{ and } \text{prob}(W / U) = \zeta \\ \text{and } \text{prob}(X / U) = \alpha \text{ and } \text{prob}(Y / U) = \beta \text{ and } \text{prob}(Z / U) = \gamma \end{array} \right) \\ &= \text{prob}_{X,Y,Z} \left( \begin{array}{l} \text{prob}(X \& Y / Z) \approx \frac{r \cdot s}{\delta} / \\ (\exists W)(\exists g)(\exists \zeta)[X, Y, Z, W \leq U \text{ and } \text{prob}(X / Z) = r \text{ and } \text{prob}(Y / Z) = s \text{ and } \\ (X \& Z) \leq W \text{ and } (Y \& Z) \leq \text{ and } \text{prob}(W / Z) = g \text{ and } \text{prob}(W / U) = \zeta \\ \text{and } \text{prob}(X / U) = \alpha \text{ and } \text{prob}(Y / U) = \beta \text{ and } \text{prob}(Z / U) = \gamma] \end{array} \right) \end{aligned}$$

and hence merely knowing that  $(\exists D)[(A \& C) \leq D \text{ and } (B \& C) \leq D \text{ and } \text{prob}(D/C) \neq 1]$  does not give us a defeater. In fact, it is a theorem of the calculus of nomic probabilities that if  $\Box[B \rightarrow C]$  then  $\text{prob}(A/B) = \text{prob}(A/B \& C)$ . So because

$$\begin{aligned} & \Box[(\text{prob}(A/C) = r \text{ and } r, s < 1 \text{ and } \text{prob}(B/C) = s) \\ & \rightarrow (\exists D)(\exists g)(\exists \zeta)[(A \& C) \leq D \text{ and } (B \& C) \leq D \text{ and } \text{prob}(D/C) = g \text{ and } \text{prob}(D/U) = \zeta]] \end{aligned}$$

it follows that

$$\text{prob}_{X,Y,Z} \left( \begin{array}{l} \text{prob}(X \& Y / Z) \approx \frac{r \cdot s}{\delta} / \\ (\exists W)(\exists g)(\exists \zeta)[X, Y, Z, W \leq U \text{ and } \text{prob}(X / Z) = r \text{ and } \text{prob}(Y / Z) = s \text{ and } \\ (X \& Z) \leq W \text{ and } (Y \& Z) \leq \text{ and } \text{prob}(W / Z) = g \text{ and } \text{prob}(W / U) = \zeta \\ \text{and } \text{prob}(X / U) = \alpha \text{ and } \text{prob}(Y / U) = \beta \text{ and } \text{prob}(Z / U) = \gamma] \end{array} \right) = 0.$$

Hence the mere fact that there always *is such a D* does not automatically give us a defeater for the application of the Principle of Statistical Independence. To get defeat, we must know of some *specific D* such that  $(A \& C) \leq D$  and  $(B \& C) \leq D$  and  $\text{prob}(D/C) \neq 1$ .

In general, defeaters for expectable probability principles are defeaters for the instance of the statistical syllogism from which they are derived. These typically derive from conflicting expectable probabilities that take account of additional information. Another defeater is:

**Subproperty Defeat for Statistical Independence:**

$\lceil (B \& C) \leq D \leq C \text{ and } \text{prob}(A/D) = p \neq r \rceil$  is an undercutting defeater for the inference by the principle of statistical independence from  $\lceil \text{prob}(A/C) = r \ \& \ \text{prob}(B/C) = s \rceil$  to  $\lceil \text{prob}(A \& B/C) = r \cdot s \rceil$ .

A principle that is actually equivalent (see Pollock 2008) to the principle of expectable statistical independence is:

**Nonclassical Direct Inference:**

If  $\text{prob}(A/B) = r$ , the expectable value of  $\text{prob}(A/B \& C) = r$ .

The principle of nonclassical direct inference supports many defeasible inferences that seem intuitively reasonable but are not licensed by the probability calculus. For example, suppose we know that the probability of a twenty year old male driver in Maryland having an auto accident over the course of a year is .07. If we add that his girlfriend's name is "Martha", we do not expect this to alter the probability. There is no way to justify this assumption within a traditional probability framework, but it is justified by nonclassical direct inference.

Nonclassical direct inference should not be confused with ordinary ("classical") direct inference. The latter licenses inferences from generic probabilities to singular probabilities, whereas nonclassical direct inference licenses inferences from generic probabilities to other generic probabilities. Nevertheless, as I will show in the next section, the two kinds of direct inference are closely related, and nonclassical direct inference provides the foundation for classical direct inference.

Because nonclassical direct inference is equivalent to the principle of expectable statistical independence, we get corresponding defeaters:

**Subproperty Defeat for Nonclassical Direct Inference:**

$\lceil B \& C \leq D \leq B \text{ and } \text{prob}(A/D) = s \neq r \rceil$  is an undercutting defeater for the inference by nonclassical direct inference from  $\lceil \text{prob}(A/B) = r \rceil$  to  $\lceil \text{prob}(A/B \& C) = r \rceil$ .

**Overlap Defeat for Nonclassical Direct Inference:**

$\lceil A \& B \leq G, C \& B \leq G \text{ and } \text{prob}(G/B) \neq 1 \rceil$  is an undercutting defeater for the inference from  $\lceil \text{prob}(A/B) = r \rceil$  to  $\lceil \text{prob}(A/B \& C) = r \rceil$  by Nonclassical Direct Inference.

These principles will be of particular importance when we turn to classical direct inference.

## 5. Direct Inference

In direct inference, we identify the value of a singular probability with the value of an associated generic probability. The basic idea is due to Reichenbach (1949). Letting **KP** mean "*P* is known", I proposed in my (1983, 1990) that this can be captured with the following two principles:

**Classical Direct Inference:**

$\lceil \text{K}Sa_1 \dots a_n \text{ and } \text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \ \& \ Tx_1 \dots x_n) = r \rceil$  is a defeasible reason for  $\lceil \text{PROB}(Ra_1 \dots a_n / Ta_1 \dots a_n) = r \rceil$ .

### Subproperty Defeat for Classical Direct Inference:

$\lceil S \leq V, \text{KV}a_1 \dots a_n, \text{and } \text{prob}(Rx_1 \dots x_n / Vx_1 \dots x_n \& Tx_1 \dots x_n) \neq r \rceil$  is an undercutting defeater for the inference by classical direct inference from  $\lceil Sa_1 \dots a_n \text{ is known and } \text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \& Tx_1 \dots x_n) = r \rceil$  to  $\lceil \text{PROB}(Ra_1 \dots a_n / Ta_1 \dots a_n) = r \rceil$ .

For example, suppose we know that Herman is an adult male of Slavic descent, and the generic probability of such a person being lactose intolerant is .35. By classical direct inference, we have a defeasible reason for expecting that  $\text{PROB}(\text{Herman is lactose intolerant}) = .35$ . But suppose we also know that Herman tests positive on a test for lactose intolerance, and the generic probability of an adult male of Slavic descent who tests positive on that test being lactose intolerant is .85. That gives us a subproperty defeater for the first inference, and a new defeasible reason for expecting that  $\text{PROB}(\text{Herman is lactose intolerant}) = .85$ . That is the intuitively right inference to make.

In light of the theory of probable probabilities, both the principle of classical direct inference and the principle of subproperty defeat for classical direct inference are theorems of the theory of nomic probability. They need not be independently postulated. By the representation theorem for singular probabilities,

$$\text{PROB}(Ra_1 \dots a_n / Ta_1 \dots a_n) = \text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \& x_1 = a_1 \& \dots \& x_n = a_n \& \mathbf{K}).$$

But the principle of non-classical direct inference gives us a defeasible reason for expecting that if  $\text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \& Tx_1 \dots x_n) = r$  then  $\text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \& x_1 = a_1 \& \dots \& x_n = a_n \& \mathbf{K}) = r$ . Thus we get the principle of classical direct inference. And the principle of subproperty defeat for classical direct inference then becomes the same thing as subproperty defeat for this instance of non-classical direct inference.

In my (1990), I suggested that a full theory of direct inference requires some additional defeaters over and above subproperty defeat. Kyburg (1974) made similar observations. I will explore some of these in the next section, but first note that from overlap defeat for non-classical direct inference, we get a defeater for classical direct inference that has previously been overlooked in the literature on direct inference:

### Overlap Defeat for Classical Direct Inference:

The conjunction of

- (i)  $Rx_1 \dots x_n \& Sx_1 \dots x_n \& Tx_1 \dots x_n \leq Gx_1 \dots x_n$  and
- (ii)  $\mathbf{K}(Sa_1 \dots a_n \& Ta_1 \dots a_n \& Ga_1 \dots a_n)$  and
- (iii)  $\text{prob}(Gx_1 \dots x_n / Sx_1 \dots x_n \& Tx_1 \dots x_n) \neq 1$

is an undercutting defeater for the inference by classical direct inference from  $\lceil \mathbf{K}Sa_1 \dots a_n$  and  $\text{prob}(Rx_1 \dots x_n / Sx_1 \dots x_n \& Tx_1 \dots x_n) = r \rceil$  to  $\lceil \text{PROB}(Ra_1 \dots a_n / Ta_1 \dots a_n) = r \rceil$ .

Because singular probabilities are generic probabilities in disguise, we can also use nonclassical direct inference to infer singular probabilities from singular probabilities. Thus  $\lceil \text{PROB}(P/Q) = r \rceil$  gives us a defeasible reason for expecting that  $\text{PROB}(P/Q\&R) = r$ . We can employ principles of statistical independence similarly. For example,  $\lceil \text{PROB}(P/R) = r \& \text{PROB}(Q/R) = s \rceil$  gives us a defeasible reason for expecting that  $\text{PROB}(P\&Q/R) = r \cdot s$ .

Thus the theory of direct inference gets a firm mathematical foundation. Nothing has to be independently postulated. It consists entirely of a set of theorems derived from the calculus of nomic probabilities and the statistical syllogism. As we will see in the next section, however, there are some important complications that remain to be explored.

## 6. The Y-Function

Although direct inference is occasionally useful, very often it cannot be used because we know too much. Suppose we have two seemingly unrelated diagnostic tests for a disease, and Bernard tests positive on both tests. We know that the probability of his having the disease if he tests

positive on the first test is .8, and the probability if he tests positive on the second test is .75. But what should we conclude about the *joint probability* of his having the disease if he tests positive on both tests? The probability calculus gives us no guidance here. It is consistent with the probability calculus for the joint probability to be anything from 0 to 1. Nor does direct inference help. Direct inference gives us one reason for thinking that the probability of Bernard having the disease is .8, and it gives us a different reason for drawing the conflicting conclusion that the probability is .75. It gives us no way to combine the information. Intuitively, it seems that the probability of his having the disease should be higher if he tests positive on both tests. But how can we justify this?

This is a general problem for theories of both classical and nonclassical direct inference. When we have some conjunction  $\lceil G_1 \& \dots \& G_n \rceil$  of properties and we want to know the value of  $\text{prob}(F/G_1 \& \dots \& G_n)$ , if we know that  $\text{prob}(F/G_1) = r$  and we don't know anything else of relevance, by nonclassical direct inference we can infer defeasibly that  $\text{prob}(F/G_1 \& \dots \& G_n) = r$ . Similarly, if we know that an object  $a$  has the properties  $G_1, \dots, G_n$  and we know that  $\text{prob}(F/G_1) = r$  and we don't know anything else of relevance, by classical direct inference we can infer defeasibly that  $\text{PROB}(Fa) = r$ . The difficulty is that we usually know more. We typically know the value of  $\text{prob}(F/G_i)$  for some  $i \neq 1$ . If  $\text{prob}(F/G_i) = s \neq r$ , we have defeasible reasons for both  $\lceil \text{prob}(F/G_1 \& \dots \& G_n) = r \rceil$  and  $\lceil \text{prob}(F/G_1 \& \dots \& G_n) = s \rceil$ , and also for both  $\lceil \text{PROB}(Fa) = r \rceil$  and  $\lceil \text{PROB}(Fa) = s \rceil$ . As these conclusions are incompatible they all undergo collective defeat. Thus the standard theory of direct inference leaves us without a conclusion to draw.

Knowledge of generic probabilities would be vastly more useful in real application if there were a function  $Y(r,s)$  such that, in a case like the above, when  $\text{prob}(F/G) = r$  and  $\text{prob}(F/H) = s$ , we could defeasibly expect that  $\text{prob}(F/G\&H) = Y(r,s)$ , and hence (by nonclassical direct inference) that  $\text{PROB}(Fa) = Y(r,s)$ . We might call this *computational inheritance*, because it computes a new value for  $\text{PROB}(Fa)$  from previously known generic probabilities. Classical direct inference, by contrast, is a kind of "noncomputational inheritance". It is *direct* in that  $\text{PROB}(Fa)$  simply inherits a value from a known generic probability. Although this might not seem entirely appropriate, I will use the generic term "direct inference" to include both classical direct inference and computational inheritance, and any other inferences from generic probabilities to singular probabilities. I will call the function used in computational inheritance "the Y-function" because its behavior would be as diagrammed in figure 1.

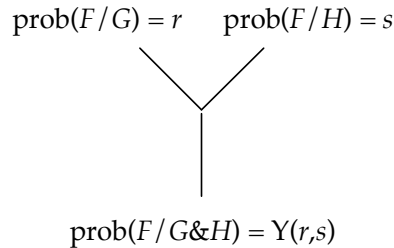


Figure 1. The Y-function

It has generally been assumed that there is no such function as the Y-function (Reichenbach 1949). Certainly, there is no function  $Y(r,s)$  such that we can conclude *deductively* that if  $\text{prob}(F/G) = r$  and  $\text{prob}(F/H) = s$  then  $\text{prob}(F/G\&H) = Y(r,s)$ . For any  $r$  and  $s$  that are neither 0 nor 1,  $\text{prob}(F/G\&H)$  can take any value between 0 and 1. However, that is equally true for nonclassical direct inference. That is, if  $\text{prob}(F/G) = r$  we cannot conclude deductively that  $\text{prob}(F/G\&H) = r$ . Nevertheless, that will tend to be the case, and we can defeasibly expect it to be the case. Might something similar be true of the Y-function? That is, could there be a function  $Y(r,s)$  such that we can defeasibly expect  $\text{prob}(F/G\&H)$  to be  $Y(r,s)$ ? It follows from the Probable Probabilities Theorem that the answer is "Yes". It is more useful to begin by looking at a three-place function rather than a two-place function. Let us define:

$$Y(r,s|a) = \frac{rs(1-a)}{a(1-r-s) + rs}$$

I use the non-standard notation “ $Y(r,s|a)$ ” rather than “ $Y(r,s,a)$ ” because the first two variables will turn out to work differently than the last variable.

The probable probabilities theorem implies that there are expectable values for joint probabilities, and the following theorem (proven in Pollock 2008, and also provable automatically using the probable probabilities software) shows what they are:

**Y-Principle:**

If  $B, C \leq U$ ,  $\text{prob}(A/B) = r$ ,  $\text{prob}(A/C) = s$ , and  $\text{prob}(A/U) = a$ , then the expectable value of  $\text{prob}(A/B \& C) = Y(r,s|a)$ , and  $B$  and  $C$  are expectably independent of both  $A$  and  $\sim A$ .

To get a better feel for what the principle of computational inheritance tells us, it is useful to examine plots of the Y-function. Figure 2 illustrates that  $Y(r,s|.5)$  is symmetric around the right-leaning diagonal.

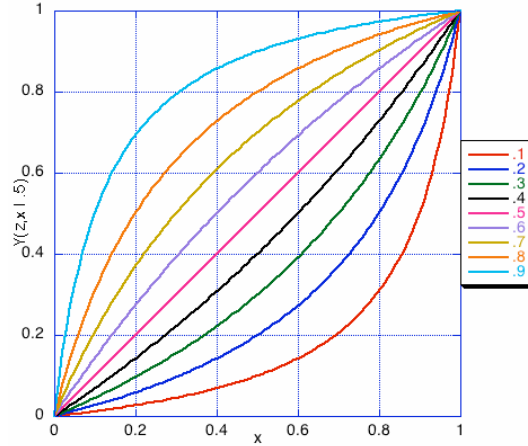
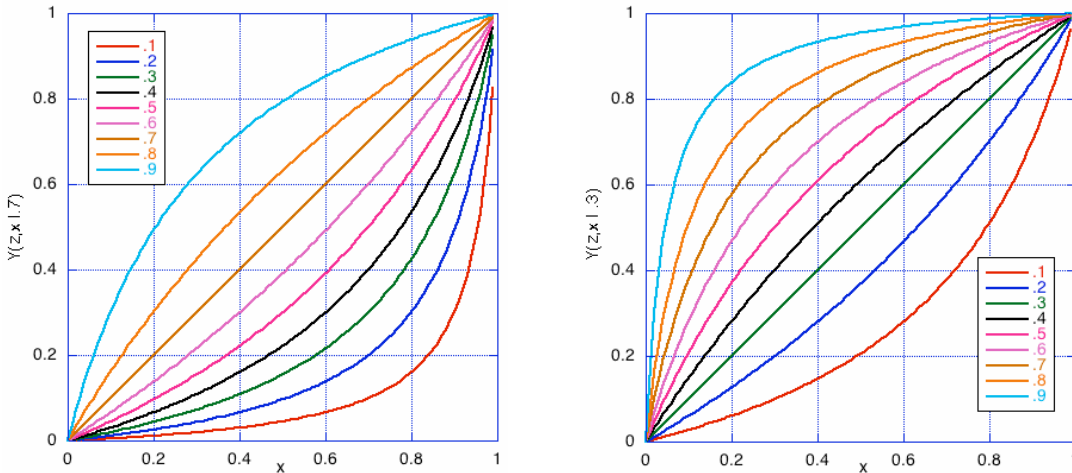


Figure 2.  $Y(z,x|.5)$ , holding  $z$  constant (for several choices of  $z$  as indicated in the key).

Varying  $a$  has the effect of warping the Y-function up or down relative to the right-leaning diagonal. This is illustrated in figure 3 for several choices of  $a$ .



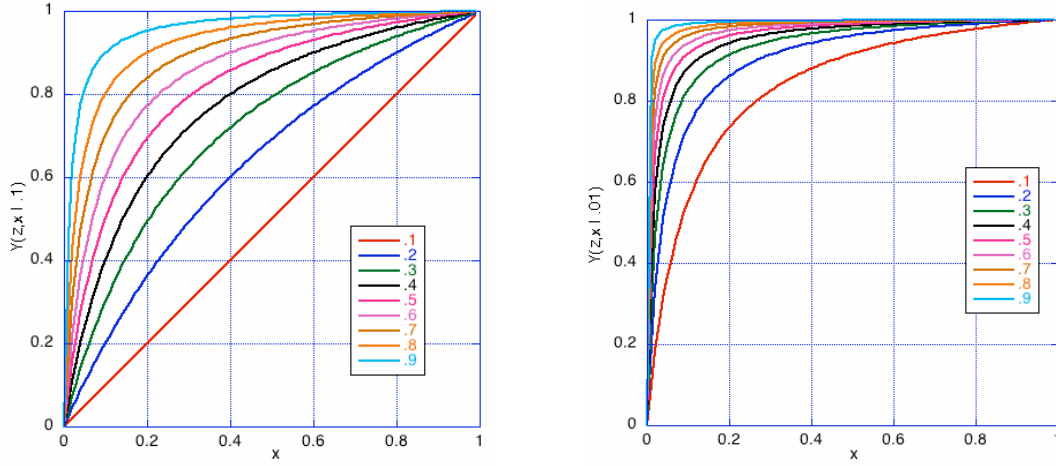


Figure 3.  $Y(z, x | a)$  holding  $z$  constant (for several choices of  $z$ ), for  $a = .7$ ,  $a = .3$ ,  $a = .1$ , and  $a = .01$ .

The Y-function has a number of important properties. In particular, it is important that the Y-function is commutative and associative in the first two variables:

**Theorem 1:**  $Y(r, s | a) = Y(s, r | a)$ .

**Theorem 2:**  $Y(r, Y(s, t | a) | a) = Y(Y(r, s | a), t | a)$ .

Theorems 1 and 2 are very important for the use of the Y-function in computing probabilities. Suppose we know that  $\text{prob}(A/B) = .6$ ,  $\text{prob}(A/C) = .7$ , and  $\text{prob}(A/D) = .75$ , where  $B, C, D \leq U$  and  $\text{prob}(A/U) = .3$ . In light of theorems 1 and 2 we can combine the first three probabilities in any order and infer defeasibly that  $\text{prob}(A/B \& C \& D) = Y(.6, Y(.7, .75 | .3) | .3) = Y(Y(.6, .7 | .3), .75 | .3) = .98$ . This makes it convenient to extend the Y-function recursively so that it can be applied to an arbitrary number of arguments (greater than or equal to 3):

If  $n \geq 3$ ,  $Y(r_1, \dots, r_n | a) = Y(r_1, Y(r_2, \dots, r_n | a) | a)$ .

Then we can then strengthen the Y-Principle as follows:

**Compound Y-Principle:**

If  $B_1, \dots, B_n \leq U$ ,  $\text{prob}(A/B_1) = r_1, \dots, \text{prob}(A/B_n) = r_n$ , and  $\text{prob}(A/U) = a$ , the expectable value of  $\text{prob}(A/B_1 \& \dots \& B_n \& C) = Y(r_1, \dots, r_n | a)$ .

If we know that  $\text{prob}(A/B) = r$  and  $\text{prob}(A/C) = s$ , we can also use nonclassical direct inference to infer defeasibly that  $\text{prob}(A/B \& C) = r$ . If  $s \neq a$ ,  $Y(r, s | a) \neq r$ , so this conflicts with the conclusion that  $\text{prob}(A/B \& C) = Y(r, s | a)$ . However, as above, the inference described by the Y-principle is based upon a probability with a more inclusive reference property than that underlying Nonclassical Direct Inference (that is, it takes account of more information), so it takes precedence and yields an undercutting defeater for Nonclassical Direct Inference:

**Y-Defeat for Nonclassical Direct Inference:**

“ $A, B, C \leq U$  and  $\text{prob}(A/C) \neq \text{prob}(A/U)$ ” is an undercutting defeater for the inference from “ $\text{prob}(A/B) = r$ ” to “ $\text{prob}(A/B \& C) = r$ ” by Nonclassical Direct Inference.

It follows that we also have defeater for the principle of statistical independence:

### Y-Defeat for Statistical Independence:

“ $A, B, C \leq U$  and  $\text{prob}(A/B) \neq \text{prob}(A/U)$ ” is an undercutting defeater for the inference from

“ $\text{prob}(A/B) = r$  &  $\text{prob}(A/C) = s$ ” to “ $\text{prob}(A \& B/C) = r \cdot s$ ” by Statistical Independence.

The Y-principle makes knowledge of generic probabilities useful in ways it was never previously useful. It tells us how to combine different probabilities that would lead to conflicting direct inferences and still arrive at a univocal value. Consider Bernard again, who has symptoms suggesting a particular disease, and tests positive on two independent tests for the disease. Suppose the probability of a person with those symptoms having the disease is .6. Suppose the probability of such a person having the disease if they test positive on the first test is .7, and the probability of their having the disease if they test positive on the second test is .75. What is the probability of their having the disease if they test positive on both tests? We can infer defeasibly that it is  $Y(.7, .75 | .6) = .875$ . We can then apply classical direct inference to conclude that the probability of Bernard’s having the disease is .875. This is a result that we could not have gotten from the probability calculus alone or from classical direct inference alone. Similar reasoning will have significant practical applications.

Joseph Halpern has pointed out to me (in correspondence) that in the special case in which the base rate is .5, the Y-principle is equivalent to Dempster’s “rule of composition” for belief functions (Shafer 1976).<sup>11</sup> However, by ignoring the base rate  $\text{prob}(A/U)$  (setting it equal to .5 by default), the Dempster-Shafer theory will often give intuitively incorrect results. For example, in the case of the two tests for the disease, two positive tests should increase that probability. But let us change the probabilities in the example and suppose that the base rate is .3, and each positive test individually confers a probability of .4 that the patient has the disease. Two positive tests should increase that probability further. Indeed,  $Y(.4, .4 | .3) = .5$ . However,  $Y(.4, .4 | .5) = .3$ , so if we ignore the base rate, two positive tests would lower the probability of having the disease instead of raising it.

Again, because singular probabilities are generic probabilities in disguise, we can apply computational inheritance to them as well and infer defeasibly that if  $\text{PROB}(P) = a$ ,  $\text{PROB}(P/Q) = r$ , and  $\text{PROB}(P/R) = s$  then  $\text{PROB}(P/Q \& R) = Y(r, s | a)$ .

The Y-Principle itself has defeaters. Let us define:

*B and C are Y-independent for A relative to U iff  $A, B, C \leq U$  and*

(a)  $\text{prob}(C/B \& A) = \text{prob}(C/A)$

and

(b)  $\text{prob}(C/B \& \sim A) = \text{prob}(C/U \& \sim A)$ .

The Y-principle derives from a defeasible assumption of Y-independence. The following theorem was proven in Pollock (2008):

### Y-Theorem:

Let  $r = \text{prob}(A/B)$ ,  $s = \text{prob}(A/C)$ , and  $a = \text{prob}(A/U)$ . If B and C are Y-independent for A relative to U then  $\text{prob}(A/B \& C) = Y(r, s | a)$ .

Somewhat surprisingly, it follows from the probability calculus that when  $\text{prob}(C/A) \neq \text{prob}(C/U)$  and  $\text{prob}(B/A) \neq \text{prob}(B/U)$ , Y-independence conflicts with ordinary independence (Pollock 2008):

**Theorem 3:** If B and C are Y-independent for A relative to U and  $\text{prob}(C/A) \neq \text{prob}(C/U)$  and  $\text{prob}(B/A) \neq \text{prob}(B/U)$  then  $\text{prob}(C/B) \neq \text{prob}(C/U)$ .

Theorem 3 seems initially surprising, because we have an initial defeasible assumption of independence for B and C relative to all three of A,  $U \& \sim A$ , and U. Theorem 3 tells us that if A is statistically relevant to B and C then we cannot have all three. However, this situation is

<sup>11</sup> See also Bacchus et al (1996). Given very restrictive assumptions, their theory gets the special case of the Y-Principle in which  $a = .5$ , but not the general case.



commonplace. Consider the example of two sensors  $B$  and  $C$  sensing the presence of an event  $A$ . Given that one sensor fires, the probability of  $A$  is higher, but raising the probability of  $A$  will normally raise the probability of the other sensor firing. So  $B$  and  $C$  are not statistically independent relative to  $U$ . However, knowing whether an event of type  $A$  is occurring screens off the effect of the sensors on one another. For example, knowing that an event of type  $A$  occurs will raise the probability of one of the sensors firing, but knowing that the other sensor is firing will not raise that probability further. So  $\text{prob}(B/C \& A) = \text{prob}(B/A)$  and  $\text{prob}(B/C \& \sim A) = \text{prob}(B/U \& \sim A)$ .

The defeasible presumption of Y-independence for  $A$  is based upon a probability that takes account of more information than the probability grounding the defeasible presumption of statistical independence relative to  $U$ , so the former takes precedence. In other words, in light of theorem 3, we get a defeater for Statistical Independence whenever we have an  $A \leq U$  such that  $\text{prob}(A/C) \neq \text{prob}(A/U)$  and  $\text{prob}(A/B) \neq \text{prob}(A/U)$ :

#### Y-Defeat for Statistical Independence:

$\lceil \text{prob}(A/C) \neq \text{prob}(A/U) \text{ and } \text{prob}(A/B) \neq \text{prob}(A/U) \rceil$  is an undercutting defeater for the inference from  $\lceil \text{prob}(A/C) = r \text{ and } \text{prob}(B/C) = s \rceil$  to  $\lceil \text{prob}(A \& B/C) = r \cdot s \rceil$  by the Principle of Statistical Independence.

## 7. An Example

The mathematics underlying the Y-Principle is compelling. It makes it reasonable to expect that joint probabilities can usually be computed using the Y-function. However, although this is a reasonable expectation, it is not logically guaranteed to yield the right answer. So it seems wise to test this prediction on a concrete example. To generate a simple database I could use for this purpose, I pressed some of my colleagues and students into service as detectors. I generated 100 paragraphs each consisting of a list of 100 words. The word “box” occurred in 40 of the paragraphs, and the word “cat” appeared in 35. A subject was shown the 100 paragraphs sequentially, and given 5 seconds to scan each paragraph, looking for the words “box” and “cat”. I used subjects as detectors *for the absence* of these words. If a subject fails to see either word, there will be some probability that the word is absent from the paragraph, and we can estimate that probability on the basis of the data we collect. So for each subject  $S$  (each detector), we can estimate  $\text{prob}(\text{not-cat}/\text{not-}S\text{-see-cat})$  and  $\text{prob}(\text{not-box}/\text{not-}S\text{-see-box})$ .

Now consider pairs of subjects (pairs of detectors)  $S_1$  and  $S_2$ . All subjects see the same paragraphs, so we can also use the data to estimate the joint probabilities  $\text{prob}(\text{not-cat}/\text{not-}S_1\text{-see-cat} \& \text{not-}S_2\text{-see-cat})$  and  $\text{prob}(\text{not-box}/\text{not-}S_1\text{-see-box} \& \text{not-}S_2\text{-see-box})$  that a word is absent given that neither subject sees it. We know the base rates of the absence of “box” and “cat”, viz., .6 and .65. So, for each pair of subjects  $S_1$  and  $S_2$ , we can compare the measured joint probability

$\text{prob}(\text{not-cat}/\text{not-}S_1\text{-see-cat} \& \text{not-}S_2\text{-see-cat})$

with the predicted joint probability

$Y(\text{prob}(\text{not-cat}/\text{not-}S_1\text{-see-cat}), \text{prob}(\text{not-cat}/\text{not-}S_2\text{-see-cat}) \mid .65),$

and we can compare the measured joint probability

$\text{prob}(\text{not-box}/\text{not-}S_1\text{-see-box} \& \text{not-}S_2\text{-see-box})$

with the predicted joint probability

$Y(\text{prob}(\text{not-box}/\text{not-}S_1\text{-see-box}), \text{prob}(\text{not-box}/\text{not-}S_2\text{-see-box}) \mid .6).$

I used 13 subjects, which generates 78 pairs of subjects and two data points for each pair of subjects (one for “cat” and one for “box”). The data points consist of measured relative frequencies, from which we can estimate the corresponding probabilities. The mean of the ratio of the measured joint relative frequency to the joint probability predicted by the Y-function was 0.995, with a mean deviation of 0.013. I take this as strong confirmation of the correctness of the use of the Y-Principle to estimate the joint probabilities in this example.

## 8. The Generalized Y-Theorem

A slight generalization of the proof of the Y-Theorem produces:

### Generalized-Y-Theorem:

Suppose  $A, B, C \leq U$ . Let  $r = \text{prob}(A/B)$ ,  $s = \text{prob}(A/C)$ ,  $a = \text{prob}(A/U)$ ,  $\alpha = \frac{\text{prob}(C/B \& A)}{\text{prob}(C/A)}$ , and

$$\beta = \frac{\text{prob}(C/B \& \sim A)}{\text{prob}(C/U \& \sim A)}. \text{ Then } \text{prob}(A/B \& C) = \frac{1}{1 + \frac{\beta a(1-r-s+rs)}{\alpha(1-a)rs}}.$$

Let us define:

$$\text{GY}(r, s | a, \alpha, \beta) = \frac{1}{1 + \frac{\beta a(1-r-s+rs)}{\alpha(1-a)rs}}.$$

Trivially:

$$Y(r, s | a) = \frac{1}{1 + \frac{a(1-r-s+rs)}{(1-a)rs}}.$$

Thus  $\text{prob}(A/B \& C) = Y(r, s | a)$  iff  $\alpha = \beta$ .

In general,  $\text{GY}(r, s | a, \alpha, \beta) > Y(r, s | a)$  iff  $\beta < \alpha$ . Therefore:

### Y-Relevance-Theorem:

Suppose  $A, B, C \leq U$ . Let  $r = \text{prob}(A/B)$ ,  $s = \text{prob}(A/C)$ ,  $a = \text{prob}(A/U)$ ,  $\alpha = \frac{\text{prob}(C/B \& A)}{\text{prob}(C/A)}$ , and

$$\beta = \frac{\text{prob}(C/B \& \sim A)}{\text{prob}(C/U \& \sim A)}. \text{ Then:}$$

- (a) If  $\alpha < 1$  and  $\beta \geq 1$  then  $\text{prob}(A/B \& C) < Y(r, s | a)$ ;
- (b) If  $\alpha > 1$  and  $\beta \leq 1$  then  $\text{prob}(A/B \& C) > Y(r, s | a)$ ;
- (c) If  $\alpha \geq 1$  and  $\beta < 1$  then  $\text{prob}(A/B \& C) > Y(r, s | a)$ .

The Y-Principle derives from the fact that, given that  $r = \text{prob}(A/B)$ ,  $s = \text{prob}(A/C)$ , and  $a = \text{prob}(A/U)$ , the expectable values of  $\alpha$  and  $\beta$  are both 1. Thus if we have a reason for believing that one of  $\alpha$  or  $\beta$  is not equal to 1, but know nothing about the other, this gives us a reason for expecting that  $\text{prob}(A/B \& C) \neq Y(r, s | a)$ , and so constitutes a defeater for the use of the Y-Principle. We get defeaters corresponding to each of cases (a), (b), and (c) of the Y-Relevance-Theorem.

In case (a),  $B$  is “negatively Y-relevant to  $C$  for  $A$ ”, and  $B$  is not positively Y-relevant to  $C$  for  $\sim A$ . In this case,  $\text{prob}(A/B \& C)$  cannot be computed as  $Y(r, s | a)$ .  $Y(r, s | a)$  provides only an upper bound on  $\text{prob}(A/B \& C)$ . To illustrate this, let us distinguish between cases in which  $\text{prob}(A/B)$  and  $\text{prob}(A/C)$  reflect informational connections and cases in which they reflect causal connections. Examples of informational connections include diagnostic relations in medicine, sensors sensing remote events, and so forth. In these cases,  $B$  does not cause  $A$  (e.g., the results of the test do not cause the disease). Rather, the direction of causation is from  $A$  to  $B$  (the disease causes the test to have certain results). If the connections are informational then it is usually reasonable to expect Y-

independence. But in causal cases, this is not a reasonable expectation. Suppose  $B$  and  $C$  both have a tendency to cause  $A$ . For example, poisoning a person and shooting him may both have a tendency to cause his death. In this case, we should not expect  $Y$ -independence to hold. Knowing that a person dies presumably raises the probability of his having been poisoned, but if we know that he was shot and died, this would raise the probability of his having been poisoned to a lesser degree (if at all). In general, if  $B$  and  $C$  are probabilistic causes of  $A$ , we would expect that  $\text{prob}(C/B \& A) < \text{prob}(C/A)$ . On the other hand, it still seems that we should expect that  $\text{prob}(C/B \& \sim A) = \text{prob}(C/U \& \sim A)$ . For instance, a person's not dying lowers the probability that he was poisoned, and it seems to do so just as much if we know he was shot. So this is a case of type (a), and all we can reasonably expect is that  $\text{prob}(A/B \& C) < Y(r, s | a)$ .

Informational cases are sometimes of types (b) or (c). For instance, suppose we have two sensors  $B$  and  $C$  detecting remote events of type  $A$ . But suppose there are two subtypes of events of type  $A$  — types  $A_1$  and  $A_2$  — where events of type  $A_1$  are more easily detected by the sensors than are events of type  $A_2$ . Then if one sensor detects an event, that raises the probability that it is of type  $A_1$ , and so raises the probability that the other sensor will also detect it. In other words,  $\alpha > 1$ . Similarly, there could be a class of cases in which the sensors are more likely to register false positives. In that case,  $\beta < 1$ .

In all of these latter cases, we should not expect  $Y$ -independence, and so should not use the  $Y$ -Principle to estimate the value of  $\text{prob}(A/B \& C)$  directly. However, in the latter two cases we can still use the  $Y$ -Principle to estimate the value of  $\text{prob}(A/B \& C)$  indirectly. Consider the first sort of case. By the probability calculus:

$$\text{prob}(A/B \& C) = \text{prob}(A_1/B \& C) + \text{prob}(A_2/B \& C).$$

If we know the values of  $\text{prob}(A_1/B)$ ,  $\text{prob}(A_1/C)$ ,  $\text{prob}(A_2/B)$ , and  $\text{prob}(A_2/C)$ , we can use the  $Y$ -Principle to compute values for  $\text{prob}(A_1/B \& C)$  and  $\text{prob}(A_2/B \& C)$ , and then sum them to estimate  $\text{prob}(A/B \& C)$ .

Instead of having different kinds of  $A$ 's, we might have different circumstances in which the reliability of the sensors vary. For instance, if we can partition the circumstances into two subcases  $S_1$  and  $S_2$ , we can compute

$$\text{prob}(A/B \& C) = \text{prob}(A/B \& C \& S_1) \cdot \text{prob}(S_1/B \& C) + \text{prob}(A/B \& C \& S_2) \cdot \text{prob}(S_2/B \& C).$$

Then if we know the values of  $\text{prob}(A/B \& S_1)$ ,  $\text{prob}(A/C \& S_1)$ ,  $\text{prob}(A/B \& S_2)$ , and  $\text{prob}(A/C \& S_2)$ , we can use the  $Y$ -Principle to compute values for  $\text{prob}(A/B \& C \& S_1)$  and  $\text{prob}(A/B \& C \& S_2)$ , and then use those to estimate  $\text{prob}(A/B \& C)$ .

In this way, we can often restore  $Y$ -independence by dividing cases. We can also have cases in which some continuous parameter affects the detectability of an event. For instance, in the "box" and "cat" cases, we could vary the amount of time subjects have for scanning a paragraph. This kind of case can be handled similarly to the above using probability distributions.

## 9. The $Y_0$ -function

The application of the  $Y$ -function presupposes that we know the base rate  $\text{prob}(A/U)$ . But suppose we do not. Then what can we conclude about  $\text{prob}(A/B \& C)$ ? I originally suspected that we could assume by default that  $\text{prob}(A/U) = .5$ , and so conclude that  $\text{prob}(A/B \& C) = Y(r, s | .5)$ . That would be interesting because, as remarked above, this is equivalent to Dempster's "rule of composition" for belief functions (Shafer 1976). As I illustrated in section six, if we know the value of  $\text{prob}(A/U)$  but ignore it, we will often get intuitively incorrect results by using the Dempster-Shafer rule. But if we do not know the value of  $\text{prob}(A/U)$ , can we assume it is .5? It turns out that even when we are ignorant of the base rate, the Dempster-Shafer rule does not give quite the right answer. The difficulty is that knowing the values of  $\text{prob}(A/B)$  and  $\text{prob}(A/C)$  affects the expectable value of  $\text{prob}(A/U)$ . Let us define  $Y_0(r, s)$  to be  $Y(r, s | a)$  where  $a$  is the solution to the following set of three simultaneous equations (for variable  $a$ ,  $b$ , and  $c$ , and fixed  $r$  and  $s$ ):

$$\begin{aligned} &2a^3 - (b + c - 2b \cdot r - 2c \cdot s - 3)a^2 \\ &+ (b \cdot c + 2b \cdot r - b \cdot cr + 2c \cdot s - b \cdot c \cdot s + 2b \cdot c \cdot r \cdot s - b - c + 1)a - b \cdot c \cdot r \cdot s = 0; \end{aligned}$$

$$\left( \frac{1-s}{1+(s-a)c} \right)^{1-s} \left( \frac{s}{a-s \cdot c} \right)^s = 1;$$

$$\left( \frac{1-r}{1+(r-a)b} \right)^{1-r} \left( \frac{r}{a-r \cdot b} \right)^r = 1.$$

Then we have the following principle (Pollock 2008):

**Y<sub>0</sub>-Principle:**

If  $\text{prob}(A/B) = r$  and  $\text{prob}(A/C) = s$ , then the expectable value of  $\text{prob}(A/B \ \& \ C) = Y_0(r,s)$ .

If  $a$  is the expectable value of  $\text{prob}(A/U)$  given that  $\text{prob}(A/B) = r$  and  $\text{prob}(A/C) = s$ , then  $Y_0(r,s) = Y(r,s | a)$ . However,  $a$  does not have a simple analytic characterization.  $Y_0(r,s)$  is plotted in figure 4, and the default values of  $\text{prob}(A/U)$  are plotted in figure 5. Note how the curve for  $Y_0(r,s)$  is twisted with respect to the curve for  $Y(r,s | .5)$  (in figure 2).

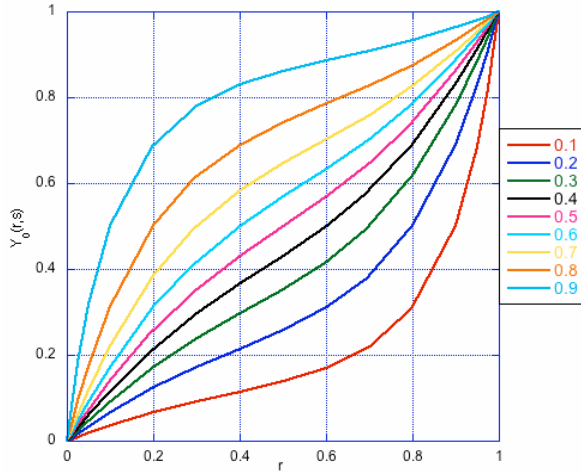


Figure 4.  $Y_0(r,s)$ , holding  $s$  constant (for several choices of  $s$  as indicated in the key)

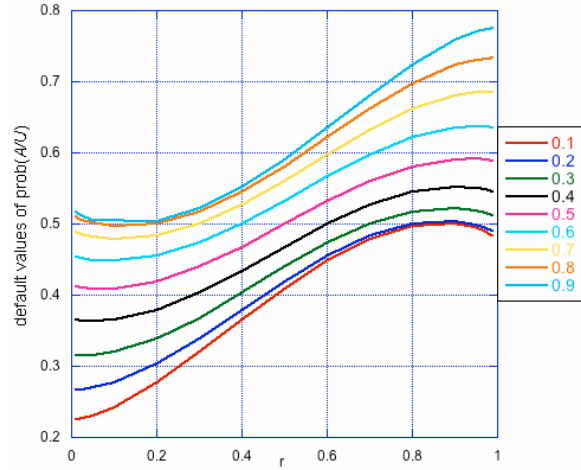


Figure 5. Default values of  $\text{prob}(A/U)$  (for several choices of  $s$  as indicated in the key)

## 10. Domination Defeat

In Pollock (1990), it was observed that there are cases in which we make direct inferences that cannot be justified solely on the basis of the principles of classical direct inference and subproperty defeat. In a statistical investigation of the cause of some kind of event (for instance, a person's getting cancer), scientists discover many factors to be irrelevant. For example, the color of one's hair is irrelevant to the probability of getting lung cancer. More precisely, the incidence of lung cancer among residents of the United States is the same as the incidence of lung cancer among redheaded residents of the United States:  $\text{prob}(C/U) = \text{prob}(C/R) = .1$ . It has also been found that the incidence of lung cancer among residents of the United States who smoke is much greater:  $\text{prob}(C/S) = .3$ . If we know that Charles is a redheaded resident of the United States who smokes, we will estimate his chances of getting lung cancer to be .3 rather than .1. But this cannot be justified in terms of classical direct inference and subproperty defeat. We have two different defeasible reasons for concluding that  $\text{PROB}(C_c) = .1$ :

- (1)  $\text{prob}(C/R) = .1 \ \& \ \text{KR}_C$

and

$$(2) \text{ prob}(C/U) = .1 \ \& \ K Uc;$$

and one defeasible reason for concluding that  $\text{PROB}(Cc) = .3$ :

$$(3) \text{ prob}(C/S) = .3 \ \& \ K Sc.$$

Because  $S \leq U$ , (3) provides a subproperty defeater for the inference from (2), but we do not have a subproperty defeater for the inference from (1). Accordingly, (1) and (3) support conflicting direct inferences and hence defeat one another, leaving us with no undefeated direct inference.

We should be able to make a direct inference to the conclusion that  $\text{prob}(Cc) = .3$ , ignoring (1). What justifies this intuitively is that we have found that being redheaded is irrelevant to the probability of getting lung cancer. Because it is irrelevant, we regard (2) as true only *because* (1) is true, and hence take any defeater for a direct inference from (1) to be a defeater for a direct inference from (2) as well. In Pollock (1990) I suggested that this can be captured as follows:

$\ulcorner \text{prob}(C/R) = \text{prob}(C/U) \ \& \ R \leq U \urcorner$  is an undercutting defeater for the direct inference from  $\text{prob}(C/R) = r \ \& \ K(Rc)$  to  $\text{prob}(Cc)$ .

I called these defeaters *domination defeaters*. Their application can be diagrammed as in figure 6, where the bold red arrow indicates a defeat relation.

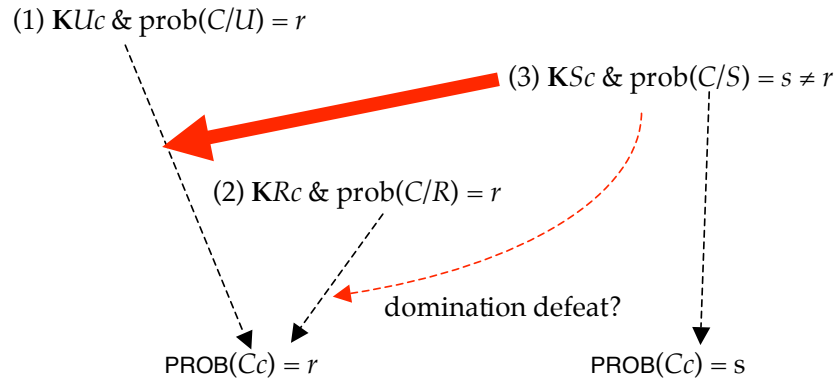


Figure 6. Domination defeat

In Pollock (1990), I suggested that defeaters of a similar structure were required for the statistical syllogism, and then domination defeaters for direct inference could be derived from those for the statistical syllogism. However, I saw no way to derive domination defeaters for the statistical syllogism from anything else, so they seemed to be a genuinely new kind of defeater that had to be postulated in addition to subproperty defeaters for the statistical syllogism. It now turns out, however, that domination defeaters for both the statistical syllogism and direct inference can be derived from the Y-principle, without the need to postulate new primitive defeaters. In Pollock (2008), I showed that domination defeaters for the statistical syllogism can be derived from the Y-principle, so domination defeaters for direct inference could be justified indirectly on that basis. However, as I will now show, they can also be derived directly from the Y-principle. This is because, by the Y-principle, if  $\text{prob}(C/U) = r$ ,  $\text{prob}(C/S) = s$ , and  $\text{prob}(C/R) = r$ , we can infer defeasibly that  $\text{prob}(C/U \ \& \ R \ \& \ S) = Y(s, r | r) = s$ , as diagrammed in figure 7. This gives us a subproperty defeater for the inference from (2) to the conclusion that  $\text{PROB}(Cc) = r$ , and leaves the inference from (3) to the conclusion that  $\text{PROB}(Cc) = s$  undefeated.

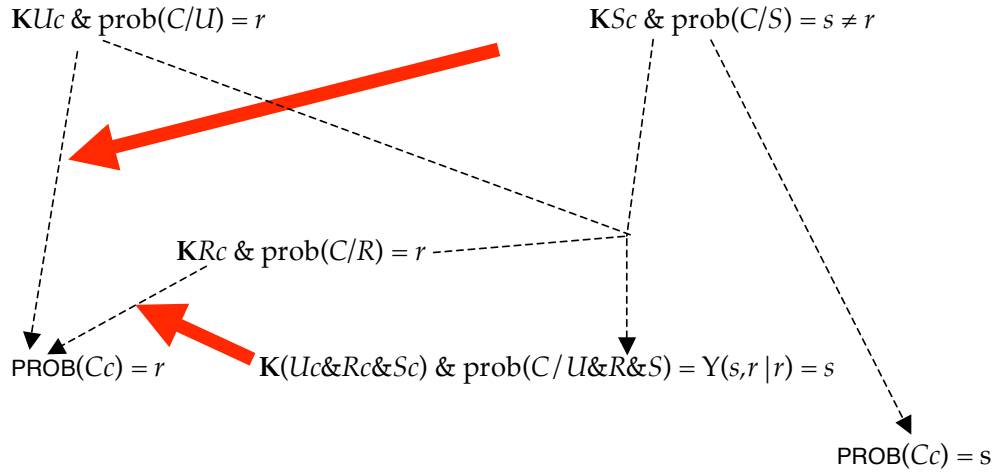


Figure 7. Reconstructing Domination Defeat

## 11. Projectibility and Disjunctive Reference Properties

In Pollock (1990), it was argued that principles of direct inference, classical direct inference, and subproperty defeat, all require projectibility constraints analogous to those required by statistical syllogism and subproperty defeat for statistical syllogism. This was defended by appeal to some well known problems for existing theories of direct inference. To get these constraints, I made some ad hoc assumptions about the projectibility of the properties to which statistical syllogism is applied in generating the principles of expectable values. However, I no longer believe that such projectibility constraints are required. In this section and the next, I will re-examine two problems that I previously took to motivate the need for projectibility constraints.

The first is “the problem of disjunctive reference properties”, due originally to Kyburg (1974), and discussed further in Pollock (1990). See also Thorn (2007). Suppose we know that  $\text{prob}(A/D) = s$ ,  $\text{prob}(A/C) = r$ , and we know that  $Bc$  and  $Dc$ , but we do not know the value of  $\text{prob}(A/B)$ . It seems that we should conclude defeasibly that  $\text{PROB}(Ac) = s$ . The value of  $\text{prob}(A/C)$  should be irrelevant, because we have no reason to believe that  $Cc$ . This seems initially to be in accordance with our theory. By the definition of singular probabilities, in this case,  $\text{PROB}(Ac) = \text{prob}(A/B \& D)$ , and by nonclassical direct inference we can infer defeasibly that  $\text{prob}(A/B \& D) = \text{prob}(A/D) = r$ .

However, although we do not know that  $Cc$ , we can infer deductively from  $Bc$  that  $(Bc \vee Cc)$ . Furthermore, although we can expect defeasibly that  $\text{prob}(A/B) = \text{prob}(A/U)$ , we would not expect that  $\text{prob}(A/B \vee C) = \text{prob}(A/U)$ . In fact, it follows from the probability calculus that if  $\text{prob}(A/C) \neq \text{prob}(A/B)$  and  $\text{prob}(C/B \vee C) \neq 0$ , then  $\text{prob}(A/B \vee C) \neq \text{prob}(A/B) = a$ . This seems to give us a subproperty defeater for the nonclassical direct inference that  $\text{prob}(A/B \& D) = \text{prob}(A/D) = s$ , and hence for the classical direct inference that  $\text{PROB}(Ac) = s$ . Because all we know about  $\text{prob}(A/B \vee C)$  follows from what we know about  $\text{prob}(A/C)$ , this should be irrelevant to the value of  $\text{prob}(A/B \& D)$ .

In Pollock (1990), I took this to illustrate the need for a projectibility constraint on both direct inference and subproperty defeat. Because projectibility is not closed under disjunction (Pollock, 1990), this would block the appeal to  $\text{prob}(A/B \vee C)$  in generating a subproperty defeater. However, it follows from the theory of probable probabilities that there is an easier solution to this problem. If we know that  $\text{prob}(A/D) = s$  and  $\text{prob}(A/C) = r$ ,  $\text{prob}(A/B \vee C)$  itself has an expectable value. In fact, running the probable probabilities software produces the following result:

```
=====
Dividing U into 4 subsets A,B,C,D, if the following constraints are satisfied:
  prob(A / C) = r
  prob(A / D) = s
  prob(A / U) = a
```

$\text{prob}(B / U) = b$   
 $\text{prob}(C / U) = c$   
 $\text{prob}(D / U) = d$

and the values of  $a, b, c, d, s, r$  are held constant, the following characterizations were found for the expectable values of the probabilities wanted:

$$\text{prob}(A / (B \vee C)) = (((r * c) + (a * b)) - (b * r * c)) / ((c + b) - (b * c))$$

$$\text{prob}(A / C) = a$$

$$\text{prob}(A / (C \& D)) = s$$

As we have seen, if  $\text{prob}(A/B \vee C) = v$ ,  $v$  will normally be different from  $\text{prob}(A/B)$ . But as long as  $v$  is the expectable value for  $\text{prob}(A/B \vee C)$ , the probability is 1 that  $\text{prob}(A/B \vee C) = v$  given that  $\text{prob}(A/C) = r$  and  $\text{prob}(A/D) = s$ , and accordingly the probability (1) of  $\text{prob}(A/B \& D)$  having any particular value given that  $\text{prob}(A/C) = r$  and  $\text{prob}(A/D) = s$  is the same as the probability (2) of  $\text{prob}(A/B \& D)$  having that value given that  $\text{prob}(A/C) = r$  and  $\text{prob}(A/D) = s$  and  $\text{prob}(A/B \vee C) = v$ . In other words, the expectable value of  $\text{prob}(A/B \& D)$  is not changed by the information that  $\text{prob}(A/B \vee C) = v$ .

What this actually shows is that the principle of subproperty defeat requires qualification. Its current formulation is:

**Subproperty Defeat for Nonclassical Direct Inference:**

$\lceil B \& C \leq D \leq B \text{ and } \text{prob}(A/D) = s \neq r \rceil$  is an undercutting defeater for the inference by nonclassical direct inference from  $\lceil \text{prob}(A/B) = r \rceil$  to  $\lceil \text{prob}(A/B \& C) = r \rceil$ .

What we see now is that this must be regarded as a defeasible defeater. Given any conjunction  $P$  of further probability statements,  $\lceil P \text{ and the expectable value of } \text{prob}(A/D) = s \text{ given } P \rceil$  is a defeater for subproperty defeat. Of course,  $P$  itself may introduce other defeaters. But in the preceding example,  $P$  is " $\text{prob}(A/C) = r$ ", and this does not introduce any new defeaters.

Subproperty defeaters for classical direct inference require a similar qualification, because they are derivable from subproperty defeaters for nonclassical direct inference.

This is a general phenomenon. Whenever we make a defeasible inference that a probability has its expectable value, this inference cannot be defeated by learning that some other probability also has its expectable value relative to some known conjunction  $P$  of additional probabilities, because the latter has probability 1. On the other hand,  $P$  itself might contain different defeaters for the inference to the expectable value. So all defeaters for principles of expectable value will be defeasible in this way.

## 12. Projectibility and Complex Target Properties

A second example that was taken in Pollock (1990) to illustrate the need for projectibility constraints concerned putative direct inferences regarding disjunctive target properties. If we know that  $\text{prob}(A \vee B/U) = r$ , running the probable probabilities software reveals, as expected, that the expectable value of  $\text{prob}(A \vee B/C) = r$ . This suggests that there is no problem applying direct inference to such properties. However, I observed that if we know that  $\text{prob}(A/C) \neq \text{prob}(A/U)$ , that seems intuitively to defeat the direct inference. This is confirmed by the theory of probable probabilities. Running the probable probabilities software produces the following results:

Dividing  $U$  into 3 subsets  $A, B, C$ , if the following constraints are satisfied:

$\text{prob}((A \vee B) / U) = r$   
 $\text{prob}(A / C) = s$   
 $\text{prob}(A / U) = a$   
 $\text{prob}(B / U) = b$   
 $\text{prob}(C / U) = c$

and the values of a, b, c, r, s are held constant, the following characterizations were found for the expectable values of the probabilities wanted:

$$\begin{aligned} \text{prob}((A \vee B) / C) &= (((s + r) - ((r * s) + a)) / (1 - a)) \\ \text{prob}(B / C) &= (((r * a) + (b * s) + (a * s)) - ((r * s) + (a^2) + (b * s * a))) / ((1 - a) * a) \end{aligned}$$

I took this to suggest that we are not really applying direct inference to the non-projectible disjunction, but instead to the projectible properties  $A$ ,  $B$ , and  $(A \& B)$ , and then inferring deductively that  $\text{prob}(A \vee B / C) = \text{prob}(A / C) + \text{prob}(B / C) - \text{prob}(A \& B)$ .

However, I now realize that this diagnosis does not solve the problem. The same problem arises for conjunctions. If we know that  $\text{prob}(A \& B / U) = r$ , the expectable value of  $\text{prob}(A \& B / C) = r$ . These is in accordance with the principle of nonclassical direct inference. However, once again, knowing that  $\text{prob}(A / C) = s \neq \text{prob}(A / U)$  defeats this inference.

Dividing  $U$  into 3 subsets  $A, B, C$  whose cardinalities relative to  $U$  are a, b, c, if the following constraints are satisfied:

$$\begin{aligned} \text{prob}(A \& B / U) &= r \\ \text{prob}(A / C) &= s \\ \text{prob}(A / U) &= a \\ \text{prob}(B / U) &= b \\ \text{prob}(C / U) &= c \end{aligned}$$

and the values of a, b, c, r, s are held constant, the following characterizations were found for the expectable values of the probabilities wanted:

$$\begin{aligned} \text{prob}(A \& B / C) &= ((r * s) / a) \\ \text{prob}(B / C) &= (((b * a) + (r * s)) - ((b * s * a) + (r * a))) / ((1 - a) * a) \end{aligned}$$

It seems to be true in general that knowing that subformulas of logically complex formulas have probabilities diverging from their expectable values changes the expectable values of the probabilities of the complex formulas. This is true for all logically complex subformulas, so a projectibility constraint is useless. This illustrates the need for a new class of defeaters for direct inference that do not follow from the defeaters adumbrated above. However, I have not proven a general theorem about this, nor do I have a general principle of defeat that I can state with confidence. This is a matter for future research.

## 13. Conclusions

Simple and intuitively unproblematic postulates regarding the proportion function that grounds nomic probability lead to a very powerful theory of probable probabilities. Combinatorial theorems about finite sets turn out to imply principles of analogous structure that license defeasible inferences about nomic probabilities. The definition of singular probabilities in terms of nomic probabilities has the consequence that analogs of these defeasible inference schemes apply to singular probabilities as well, and also license a wide variety of principles of direct inference, some not previously known. In the past, discovering what the defeaters are for principles of direct inference has proceeded in an ad hoc way, by finding examples in which direct inference yields intuitively incorrect results and then trying to construct a principle that correctly describes a class of cases in which the inferences should be blocked. This has led to some very convoluted principles with little claim of intuitive obviousness. However, within the theory of probable probabilities the discovery of defeaters reduces to the purely objective task of finding and proving combinatorial theorems about finite sets. This can often be done automatically using the probable probabilities software.



## References

- Bacchus, Fahiem  
 1990 *Representing and Reasoning with Probabilistic Knowledge*, MIT Press.
- Bacchus, Fahiem, Adam J. Grove, Joseph Y. Halpern, Daphne Koller  
 1996 "From statistical knowledge bases to degrees of belief", *Artificial Intelligence* **87**, 75-143.
- Braithwaite, R. B.  
 1953 *Scientific Explanation*. Cambridge: Cambridge University Press.
- Carnap, Rudolph  
 1950 *The Logical Foundations of Probability*. Chicago: University of Chicago Press.  
 1952 *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.
- Fisher, R. A.  
 1922 "On the mathematical foundations of theoretical statistics." *Philosophical Transactions of the Royal Society A*, 222, 309-368.
- Goodman, Nelson  
 1955 *Fact, Fiction, and Forecast*, Cambridge, Mass.: Harvard University Press.
- Halpern, J. Y.  
 1990 "An analysis of first-order logics of probability", *Artificial Intelligence* **46**, 311-350.
- Jackson, Frank, and John Pargetter  
 1973 Indefinite probability statements. *Synthese* **26**, 205-215.
- Kneale, William  
 1949 *Probability and Induction*. Oxford: Oxford University Press.
- Kyburg, Henry, Jr.  
 1961 *Probability and the Logic of Rational Belief*. Middletown, Conn.: Wesleyan University Press.  
 1974 *The Logical Foundations of Statistical Inference*, Dordrecht: Reidel.  
 1974a "Propensities and probabilities." *British Journal for the Philosophy of Science* **25**, 321-353.
- Pollock, John L.  
 1983 "A theory of direct inference", *Theory and Decision* **15**, 29-96.  
 1984 "Foundations for direct inference". *Theory and Decision* **17**, 221-256.  
 1990 *Nomic Probability and the Foundations of Induction*, New York: Oxford University Press.  
 2006 *Thinking about Acting: Logical Foundations for Rational Decision Making*, New York: Oxford University Press.  
 2008 "Probable probabilities", <http://oscarhome.soc-sci.arizona.edu/ftp/PAPERS/ProbableProbabilities.pdf>.  
 2008a "Defeasible reasoning", in *Reasoning: Studies of Human Inference and its Foundations*, (ed) Jonathan Adler and Lance Rips, Cambridge: Cambridge University Press.
- Popper, Karl  
 1956 "The propensity interpretation of probability." *British Journal for the Philosophy of Science* **10**, 25-42.  
 1957 "The propensity interpretation of the calculus of probability, and the quantum theory." In *Observation and Interpretation*, ed. S. Körner, 65-70. New York: Academic Press.  
 1959 *The Logic of Scientific Discovery*, New York: Basic Books.
- Reichenbach, Hans  
 1949 *A Theory of Probability*, Berkeley: University of California Press. (Original German edition 1935.)
- Shafer, G.  
 1976 *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
- Sklar, Lawrence  
 1970 "Is propensity a dispositional concept?" *Journal of Philosophy* **67**, 355-366.  
 1973 "Unfair to frequencies." *Journal of Philosophy* **70**, 41-52.
- Thorn, Paul  
 2007 *Three Problems of Direct Inference*, PhD dissertation, Dept. of Philosophy, University of Arizona.
- van Fraassen, Bas  
 1981 *The Scientific Image*. Oxford: Oxford University Press.
- Venn, John  
 1888 *The Logic of Chance*, 3rd ed. London.