prolific output). They have spent their lives together, they have read the same books, and are in fact *zwei Seelen und ein Gedanke.* Specifically, they share the same subjective probabilities. Give them both the same data and the same problem. The Doogian formula implies that they will come to different conclusions for two different reasons.

1. IJG may have more time to introspect upon the problem that IJG′, so that the twins settle upon different interval-valued probabilities.
2. Even if the twins decide upon the same intervals, there may be several sorts of analysis that do not contradict these beliefs; thus IJG and IJG′ could pursue different analyses and reach different conclusions.

The first problem is intrinsic to the Type II view; the second problem is what seems most awkward for the Bayes/non-Bayes compromise. One needs a basis for deciding when enough thought has given to the problem, and a protocol for choosing among equally justified analyses.

CAUSALITY AND CASUISTRY

Probabilistic causality is an elegant piece of argument, and the concept seems potentially important, but it is not yet clear that the subject will be relevant to statisticians. After rereading the expanded treatment of the subject given in Good (1983f, Chapters 21 and 22), I am left with the fear that this is a case of *molto fumo e poco arrosto.* The philosophical side of the house seems very pleased with it, but one should note the historical tendency for a subject to split off from philosophy as soon as it becomes respectable (e.g., cosmology, mathematics, decision theory, etc.).

Good's rationale for the statistical value of probabilistic causality rests on the fact that $Q(E:F)$ agrees with a measure of association in contingency tables, and that it can be used to interpret the expected influence of a change in the regressor variables in linear regression. This does establish a connection with statistics, but I look forward to future developments in this area that will be more compelling. One possible application that enjoys the advantage of topicality is a plot over time of the estimated value of $Q(E:F)$, the tendency of smoking to cause lung cancer, based on the data available to the tobacco industry in 1945, 1950, ..., 1985.

A similar comment applies to explicativity, in that it isn't clearly crucial to modern statistics. There is the possibility of important connections with model selection, and Good mentions the work of Akaike (1974) and Schwarz (1978), but the key comparisons have yet to be made. This is another area in which one hopes that Good's article will strike sparks.

ADDITIONAL REFERENCES

BERGER, J. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

HUBER, P. (1973). The use of choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45** 181–191.

Rejoinder

I. J. Good

I'm most grateful to the discussants for their comments, both the generous and the critical ones. All four discussants seem to approve of some form of Bayes/non-Bayes compromise and with some other things I've said, but they have raised various issues that demand some response. I shall respond in the order in which the contributions are printed, but I deal first with probabilistic causality because three of the discussants have commented on it.

1. PROBABILISTIC CAUSALITY

The notation $Q(E:F)$ is an abbreviated notation, and, as I mentioned, the full notation mentions "the true laws of nature" (and other things) as *given* (to the right of a vertical stroke). When there are two or more scientific theories there will therefore be two or more estimates of $Q(E:F)$, and at most one of those estimates can be correct. This is my reply to one of the comments by Suppes where he described two different theories of learning only one of which can be (approximately) true.

I agree with Suppes's analogy with regression theory, in fact it is somewhat more than an analogy. If much of some other science is taken into account in a statistical or philosophical project, then the project is no longer regarded as just statistics or just philosophy. A physicist usually wants a better explanation of his data sets than can be provided by regression theory (or by *estimated* probabilistic causal tendency) alone; but a chemical engineer, a physician, an economist or a sociologist is usually somewhat less ambitious.

One advantage of any quantitative or semiquantitative rather than a purely qualitative theory, apart of course from quantitative applications, is that it shows more clearly the self-consistency of the assumptions (desiderata) and can thus make a qualitative theory more acceptable. Also, as I said, a formula can encapsulate many words.

Barnard expressed doubts about the value of the word *cause* within science, rather than within legal matters. Bertrand Russell expressed a similar sentiment in 1913, mainly in regard to physics, and is quoted at some length by Suppes (1970, page 5) who goes on to say "Perhaps the most amusing thing about this passage from Russell is that its claim about the use of the word 'cause' in physics no longer holds." Suppes supports his assertion with irrefutable evidence as of 1970. Outside of physics, as in medical and agricultural science, it seems to me that the attempt to discover causes, as distinct from correlations alone, is usually a goal in the design of randomized statistical experiments. We are controlled by nature, but by discovering causes we can recover some of the control. I believe Barnard must have been exaggerating when he expressed doubts about the value of the word *cause* in science.

An aspect of probabilistic causality that should be investigated would be its potential use in path analysis, as understood by Sewall Wright. See, for example, Wright (1968, pages 299–372).

I agree with Banks that philosophical topics, when sufficiently well developed, tend to move into other disciplines, and I, and probably others, have made the same point elsewhere, at least orally. (Banks uses the description *respectable* instead of *sufficiently well developed.*) But my essay was about the interface between philosophy and statistics so it was appropriate to discuss some topics that have not yet crossed right through the interface. Don't forget that influences flow in both directions through the interface. I am interested in fundamental questions and surely the notions of probabilistic causality and explanation are fundamental. Check again in another fifty years. Philosophical ideas in science usually start as minority views and therefore usually take a long time to become "respectable."

I liked Banks's suggestion of plotting, as a time series, the estimated values of the tendency of smoking to cause lung cancer. It would be a large project but would be a nice example of a quantitative method in the history of science. The same idea could be applied elsewhere, such as the effects of other drugs, both "recreational" and medical.

I shall now respond to the discussants individually.

2. RESPONSE TO PATRICK SUPPES

2.1 Randomness and Complexity

I agree that one must distinguish between random procedures and random results, and that the distinction is often not made clearly enough. I made the point in 1983f; page 87 ff. (quoting Good, 1972; compare Good, 1956) when referring to suspicious-looking sequences. I cited Kendall (1941) and Scott (1958) on this topic, and recently presented a colloquium on the topic in Blacksburg. In that colloquium I gave the example of the sequence

(1) 99992128599999399

which occurs as 17 consecutive digits in the expansion of $\pi - 3$, beginning at the 19437th digit. Few statisticians would be happy to use this sequence in a random design, although most would regard the expansion of π as a satisfactory pseudorandom sequence. (It would be amusing to ask for the next digit after sequence (1) in a nontimed very high I.Q. test. The reader of this article will be nearly sure that it is not a 9, but not by appealing to the crude fallacy of the "maturity of the chances"!)

One simple description of the sequence (1) is "digit number 19437, and the next 16 digits of $\pi - 3$." Another is "*AA*21285*AA*93*A* where $A = 99$." The probability of getting at least eleven of the same digit in a flat-random sequence of 17 digits is about 7.0×10^{-7}, but we should "pay" a factor of 20,000 for special selection if the pseudorandomness of the expansion of π is being tested (because I scanned 20,000 digits to find the sequence (1)).

I don't think it is sufficient to reject (finite) sequences of lowish complexity. This method would not lead to the rejection of all sequences that get decisively rejected even by standard statistical tests.

2.2 Determinism

The theorem that Suppes quotes is indeed remarkable, as are some of the results in recent chaotics, but I don't see that those results contradict my 1972 comment which Suppes also quotes. In his example each possible infinite sequence is a possible outcome of a deterministic set-up, corresponding to initial conditions that would have to be given with infinite accuracy to predict the output. This shows that the infinite output sequence of integers does not logically determine whether the process is deterministic or indeterministic. I wonder whether it would *support* the hypothesis of indeterminism, but I suppose this question is purely theoretical.

Note that the number of possible outputs is non-countably infinite, and so is the number of possible inputs. But the number of possible inputs that can be *specified* cannot be noncountably infinite because a specification must be of finite length. (I am liberally allowing for finite but unbounded lengths of the specifications.) In an indeterministic model, involving continuous distributions, the output can be any infinite sequence even if the inputs are fully specified in finite terms, but this is not true for a deterministic model because the cardinal number of possible outputs cannot then exceed the cardinal number of possible inputs. So there is a philosophical distinction between determinism and indeterminism. But, as I said in the 1972 passage, indeterminism arises out of determinism, for all practical purposes, because infinite accuracy of specification or measurement of the initial set-up cannot be achieved in practice. I did not then know that extreme sensitivity to the initial conditions occurs even for very simple nonlinear deterministic physical models as is familiar to chaoticians.

3. RESPONSE TO GEORGE BARNARD

3.1 Maximum Likelihood

A maximum likelihood estimate may or may not be associated with an estimate of spread. It is hardly a mistake to use a point maximum likelihood estimate on *some* occasions. I recall Barnard saying, some time in the 50s, that he suspected that most of the differences between statisticians might turn out to be semantic. That might be the case here. When there are r successes in n independent and identically distributed (iid) trials, people do sometimes regard r/n as the appropriate betting probability for the next trial, and this is the point maximum likelihood estimate. I believe that, if your state of information borders on ignorance, then even Laplace's point estimate $(r + 1)/(n + 2)$, although not ideal, is more reliable than r/n. Perhaps I should have said that no sensible person would bet on the basis of the point maximum likelihood estimate when $r = 0$ or $r = n$, in the sense of giving "infinite" odds (a soul versus a sou), unless possibly n were extremely large.

3.2 Model Adjustment Parameters and the Bayes/non-Bayes Compromise

Barnard mentions the concept of a model adjustment parameter, and asks how it fits in with the Bayes/non-Bayes compromise. I think the adjustment parameter could be regarded as just one of the parameters of the model. But it might sometimes be expedient, partly because of the available non-Bayesian software, to assign it a prior distribution while treating the other parameters in a non-Bayesian manner. This procedure would exemplify a Bayes/non-Bayes compromise.

3.3 Summarization of Data Sets

One reason why I did not say more on this topic was that I have written an article on the philosophy of exploratory data analysis (1983e). I argued that it has Bayesian aspects although superficially it might seem to be brutally empirical. Some of that discussion would apply also to the allied topic of the summarization of data sets.

As in exploratory data analysis, some devices for the summarization of data implicitly take some physiological psychology into account, as when data is represented in abstract creature-feature space and is then projected into two dimensions. There is some discussion of such matters in Good (1983e).

I think I agree with much of Barnard's partial formalization of data summarization, and that the reporting of a likelihood function (which he has advocated in the past and I think is exemplifying here) can be regarded as a Bayes/non-Bayes compromise.

3.4 Fiducial Probability

Barnard finds it hard to believe that Fisher was misled by not using a notation for conditional probability. It is also almost unbelievable that Bronstein, a world-class chess player, left his queen *en prise* in a game against Petrosian, but he did make that beginner's blunder. Fisher, like Bronstein, was human and it is easy to be misled by an incomplete notation just as most people, including philosophers, can be misled by ill-defined or ambiguous words. (But some ambiguity is necessary to avoid having to go on talking forever, an "uncertainty principle" that has been around for decades.) *Please read carefully* my reason for making the assertion without taking foregone conclusions for granted: the argument is only one page long and is I think clearly written. After eighteen years I'm still waiting for a refutation. By mentioning Kolmogorov, Barnard makes the matter seem more complicated than necessary.

In Barnard's example of the fiducial argument the absolute probability that (λ, σ) belongs to the set $\{(\lambda, \sigma); (t_0, z_0) \in S\}$ can be computed, without assuming a prior distribution for (λ, σ), as in the definition of a confidence set (in Neyman's sense). But this probability is not in general equal to the probability *conditional* on knowing the values of $(\bar{x}, s)$ which is what the fiducial argument claims to provide. The calculation of this conditional probability, unlike the absolute probability, requires the assumption of a prior distribution for (λ, σ).

Two statements, such as $x > y$ and $y < x$, can be logically equivalent but can have different probabilities. That seems like a paradox only if you forget that the statements might be conditional on two different propositions or, expressed differently, one might be a conditional probability and the other an absolute one. When we include the "given" propositions in the notation we are not tempted into a paradox.

Mistakes always look worse when they are pinpointed. George, why not publicly capitulate on this issue?

Similarly where Fisher and Barnard emphasize that likelihood does not obey Kolmogorov's axioms I would make the point by saying that $P(A \mid B) \neq P(B \mid A)$.

3.5 Weight of Evidence

Weights of evidence need not be at all precise to be useful. I believe that when a doctor, a detective or a cat smells a rat, he, she or it behaves as if she had made an implicit judgment of a weight of evidence in the technical sense.

Barnard points out that Peirce's book *Chance, Love, and Logic* discussed the technical concept of weight of evidence. I find that Chapter 4 of that book is a reprinting of Peirce (1878) to which I had referred. Barnard says that, according to his recollection, Peirce did not mention the condition that H and not-H were initially equally probable. In brief my reply is "yes and no" and, to explain this, I shall try to clarify some aspects of Peirce's obscurely expressed article.

As Isaac Levi first pointed out to me, Peirce's article anticipated the Neyman-Pearson concept of confidence intervals though not with clarity. One reason why I did not at first see this was that Peirce used the expression *weight of evidence* to mean log-odds, which sounds close to the definition as the logarithm of a Bayes factor, and he also pointed out an additive property of his definition. (I return soon to the two interpretations of "odds".) Moreover he then applied the terminology of "weight of evidence" to show that the "conceptualistic" (Bayesian or Laplacian) point of view is not entirely without merit.

Peirce says "It is entirely in harmony with this law [Fechner's law] that the feeling of belief should be as the logarithm of the chance [that is, of the odds] ... The rule for the combination of independent concurrent arguments takes a very simple form when expressed in terms of the intensity of belief, measured in the proposed way." At this point Peirce seems to be confusing, or causing the reader to confuse, the Neyman-Pearsonian odds of "rules of inference" being right (before conditioning on the outcome of an experiment) with the Bayesian posterior odds that a hypothesis is correct. Within the Neyman-Pearson theory one is not supposed to talk about the odds of a hypothesis, whether prior or posterior. Note too that later Peirce exemplifies an "argument" by means of an *event* (discovering that a bean is black).

Peirce goes on to say "take the sum of all the feelings of belief which would be produced separately by all the arguments *pro*, subtract ... the similar sum for arguments *con*, and the remainder is the feeling of belief which we ought to have on the whole." Apparently the resultant "feeling of belief" is now to be understood as the final (posterior) log-odds *of the hypothesis*. It is here that we need to assume that the initial log-odds are 1. In fact he says later "In the conceptualistic view ... complete ignorance ... is represented by the probability ½." The condition $P(H) = ½$ did not appear in the nonconceptualistic part of Peirce's discussion, so Barnard's recollection was correct in that context.

At one point Peirce's article shows a lack of understanding of the conceptualistic position. He considers a bag containing a large number of beans each known to be either black or white, and one bean is selected at random and placed under a thimble without its color being observed. Then $2n$ random drawings are made at random with replacement. Suppose that $n + 10$ are black and $n - 10$ are white. This provides evidence that the bean under the thimble is black (hypothesis H). Peirce states that the conceptualist will incorrectly think the weight of evidence in favor of H is the same whether $n = 10$ or $n = 1000$ because the excess of blacks is 20 in both cases. He did not notice that the drawings were physically but not subjectively independent. He overlooked that the prior distribution of the physical probability has to be continually updated unless the density is a Dirac function. Or, in other terms, $W(H:F \mid E)$ need not equal $W(H:F)$!

My present view is that Peirce did not after all clearly anticipate the correct Bayesian interpretation of weight of evidence although he came close enough to it to suggest it to the mind of a Bayesian reader. Incidentally, I have recently suggested a new and very succinct justification for the Bayesian interpretation (Good, 1988d).

The reading of much of Peirce's article, most generous to him and expressed in modern terms, is that it would be wrong to put a Bayesian interpretation on a confidence interval (although it is still done). I wonder if any of his readers in 1878 understood him in this manner.

Barnard has recently informed me that Cournot (1843) anticipated the concept of confidence intervals, and that this book is to be republished as Volume 1 of Cournot's collected works under the auspices of the French Centre National de la Recherche Scientifique. Hurrah for that kind of national pride! Laplace (1820, page 281) also anticipated the concept, but I don't think he gave it much emphasis.

3.6 Scientific Induction

I think it would be a pity if the word *induction* were expunged from the dictionary even though it is used ambiguously. It is better to try to disentangle its various meanings and this attempt is consistent with Barnard's belief in the value of taxonomy of concepts. I agree that induction is surrounded by confusion but it is worth more than the paper it is printed on.

I have recently noticed (Good, 1988e) that the *form* of argument, used by Popper and Miller (1983, 1987) against probabilistic induction, an argument that does not convince me at all, can, if valid, be used to demonstrate an absurd consequence. I have pointed this out to Popper and Miller and am waiting for a reply.

3.7 *P*-values

The proposal of using standardized p-values, when the null hypothesis is a simple statistical hypothesis, is not by any means intended to solve all problems of interpretation. My thesis is only that if you must use p-values in these circumstances, as measures of evidence against the null hypothesis, then they are less misleading when standardized than when they are not. I recall Barnard's saying, from the floor at a meeting of the Research Section of the Royal Statistical Society in the '60s, that a p-value is a "primitive concept" or that it is more primitive than a Bayes factor. I think this is still the attitude of many users of statistical methods, although I recognize that Barnard, in agreement with Fisher, has shifted somewhat from this opinion. At one time in the '50s, in conversation, he used to agree with Fisher's interpretation of a very small p-value: "*Either* an exceptionally rare chance has occurred, *or* the theory of random distribution is not true" (Fisher, 1956, page 39). I don't know whether Barnard was quoting Fisher, or *vice versa.* A difficulty with this interpretation is that all events are highly improbable when described in much detail. This difficulty does not occur when sets of all possible simple likelihood ratios are used, or when Bayes factors are used, and perhaps the difficulty can also be avoided by using surprise indexes.

When a Bayes factor is vague enough, or too difficult to estimate, or when dealing with a community educated to regard p-values as clearer than Bayes factors, we may have to fall back on p-values or on surprise indexes. As Bayesian techniques become more and more developed, and when Bayesian software becomes more readily available, the fraction of applications in which one needs to use p-values may well decrease, but I don't think the need will disappear within the next hundred years unless civilization disappears first. There might be a period of a few decades during which it will be common to provide both a Bayesian and a non-Bayesian analysis of each applied problem. (For an example see Haldane and Smith, 1947.) Meanwhile, examples of the very approximate relationships between p-values and Bayes factors, although not the simple relationship that a few might expect, should soon constitute a part of an elementary education in statistical methodology. This is especially important because the popular p-value of 0.05 is usually only weak evidence against a sharp null hypothesis.

The result that Barnard ascribes to Pitman (1965) is true but does not appear there explicitly. An extreme case is when an experiment is performed that has no bearing at all on the scientific matter at issue. Then no p-value, however small, standardized or not, would be relevant, and the weight of evidence would be zero. Standardization is not intended to be a panacea.

I say more about p-values in my reply to Berger's comments.

Regarding Fisher's changing views from the '20s to the '50s, these changes might well have been caused in part by the opinions of others such as Neyman and Egon Pearson and various Bayesians or near-Bayesians. The father of statistics presumably learned something from his descendants, including Barnard.

3.8 Subjective Probability

When I said that subjective probability was the most basic kind, I have to confess to an ambiguity. I meant that subjective probability should be used even when estimating a physical probability and I had mainly in mind the contexts of ordinary statistical practice and ordinary life. I tend to believe also in the reality of physical probabilities in relation to quantum mechanics, and in their usefulness as convenient fictions in many deterministic contexts. Even in quantum mechanics a physical probability can be interpreted as a subjective one when a decision needs to be made.

4. RESPONSE TO JIM BERGER

4.1 Hierarchical Bayes

Perhaps I exaggerated by implying that all Bayesian methods will eventually be hierarchical. Presumably, for the sake of simplicity, or for the sake of Type II rationality, the hierarchy will often continue to be mowed at the lowest level, before the hierarchy grows. For example, a beta prior with specific hyperparameters might often be assumed although at the back of your mind you know that you don't believe in these exact values. If you have time you should do a sensitivity analysis at any level of cut-off.

4.2 Bayes Factors

I like Jim Berger's remark about some anti-Bayesians "recoiling in horror." The expression "weighted likelihood ratio" is perhaps used by some to capture the Bayesian concept of a Bayes factor while pretending to be non-Bayesian. Weighted likelihood ratios are closely related to the work by Crook and myself on the *strength* of a statistical test (1982). This is a weighted power function where the natural weight function is a prior distribution. Kempthorne had previously suggested an *averaged* power function. This is the anti-Bayesian way of using a "Bayes postulate" (uniformity of the prior).

4.3 *P*-values

Berger's modified standardization might improve the one I proposed, but mine is a little simpler. He agrees that my 1957 notion of using a possibly crude Bayes factor as a significance test criterion (with p-values) should be beneficial to classical testing but he questions whether p-values should ever be used. My reply is partly contained in that to Barnard's contribution.

At present it is sometimes too difficult to produce a satisfactory Bayesian model, and then one is tempted to fall back on p-values and all that. There is also the problem of dealing promptly with brain-washed clients. Perhaps one day there will be satisfactory Bayesian procedures and software for nearly all occasions.

If p-values go down the drain can confidence intervals long remain?

5. RESPONSE TO DAVID BANKS

I very much like Banks's suggestion for experimental work at the border of psychology, philosophy and statistics.

In reference to his reference to insanity, compare my one-sentence publication Good (1976b) which asked "What upper bound does the possibility of hallucination put on the evidence for any hypothesis?"

Among the many other interesting ideas in Banks's contribution I'm responding just to the question of when to stop thinking about a problem (without obtaining new empirical data). The question impinges on the closely related topics of Type II rationality and dynamic probability. It seems to me that you have to judge from time to time whether the expected utility of further thought is still positive. This is a problem that everybody has to face repeatedly; for example, Banks must have faced it when he decided to stop writing his contribution to the discussion. In a game of chess, played with a chess clock, the problem arises in an especially clear manner and quite a bit can be said in that context (Good, 1968c). In more general contexts it is difficult to give specific advice other than the need to take into account whatever springs to your mind or to your client's mind, and is judged to be sufficiently relevant. One criterion is whether your thoughts have begun to go round in circles. Then, if you are sensible you will say to yourself "For heaven's sake, make your mind up!" but you would probably be more polite to your client. Sleeping on a problem is often of value, presumably because the unconscious mind is much more powerful than is sometimes recognized. It might be more powerful in achieving rationality of Type II than of Type I, and it might use a mechanism like hypnotism for this purpose.

When the famous mathematician J. E. Littlewood had no new ideas, for a couple of days, concerning a problem he was attacking, he would postpone further work for a few months during which time he hoped to acquire some relevant new knowledge. (Oral communication from Besicovitch, about 1939.) For theoretical work this approach is probably useful in statistics, but less so for statistical consulting!

It might be possible to develop new useful suggestions to help one to decide when to stop thinking about a statistical problem. One strategy is to choose a Bayesian model and to check whether it leads to conclusions that are robust or sensitive to moderate modifications. For a recent example of this strategy see Good and Crook (1987) and the additional reference in Jim Berger's contribution to the present discussion.

I believe that the procedure of stopping thinking and calculating, in accordance with Type II rationality, is what statisticians have usually done implicitly, to some approximation, for the last 200 years. I have thought it worthwhile to make the matter more explicit. As in a theory of dynamic probability, any attempt to formalize Type II rationality requires some notations like P_t, u_t and W_t (for probabilities, utilities and weights of evidence) to convey the time t at which the judgments are made. One should think of t as representing a time interval rather than a single moment of time. Note too that coherence is usually more important in a single document than in your thoughts at a given time.

When modifying your estimates of probabilities, utilities, weights of evidence, etc., you would *sometimes* rely on the usual axioms of probability, but reexpressed to bring in the time element. For example, the product law would be expressed in a manner resembling

$$P_{t+1}(E\&F \mid H) = P_t(E \mid H)P_t(F \mid E\&H),$$

$$P_{t+1}(F \mid E\&H) = P_t(E\&F \mid H)/P_t(E \mid H),$$

or

$$P_{t+1}(E \mid H) = P_t(E\&F \mid H)/P_t(F \mid E\&H),$$

where only one of these three equations can be used at any specific time in serial processing. (By $t + 1$ I mean a time interval following the one labeled t. Bringing the time into the notation is like bringing in a symbol for your "state of mind": compare, Good, 1950, pages 2 and 3.) If the subscripts are dropped then the equality symbols should be replaced by an updating symbol such as the Algol ": =." If no thought concerning $P(E \mid F)$, nor any closely relevant probability, has occurred to you between times t and $t + 1$, then one has the default

$$P_{t+1}(E \mid F) = P_t(E \mid F).$$

With such notations I don't think inconsistencies have to arise, but they probably would do so in a highly parallel machine such as the human brain. Then either the inconsistencies are ignored, being judged to be unimportant, or some further algorithms would be needed to resolve the contradictions, such as those proposed for the combination of judgments by Lindley, Tversky and Brown (1979) and Good (1979). In pure logic, any contradiction is devastating, but, in natural human probabilistic thinking contradictions can be more or less relevant to the questions of interest to you at any time. We have amazing ability to notice the relevant and to ignore the irrelevant. It might be even more difficult to formalize this process than to mechanize it with the help of an artificial neural network.

Further work on Type II rationality, dynamic probability and on the allied topic of temporal coherence can be found in Goldstein (1985), Diaconis (1987) and Michie (1987), which introduces a concept of Type III rationality in relation to pseudognostics (artificial intelligence). These papers contain further citations.

ADDITIONAL REFERENCES

COURNOT, A. A. (1843). *Exposition de la théorie des chances et des probabilités.* Hachette, Paris.

DIACONIS, P. (1987). I. J. Good and the dynamics of subjective probability. In *Foundations and Philosophy of Probability and Statistics, an International Symposium in Honor of I. J. Good on the Occasion of His 70th Birthday, May 25–26* (K. Hinklemann, ed.). *J. Statist. Plann. Inference.* To appear.

GOLDSTEIN, M. (1985). Temporal coherence (with discussion). In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 231–248. North-Holland, Amsterdam.

GOOD, I. J. (1956). Which comes first, probability or statistics? *J. Inst. Actuaries* **82** 249–255. Reprinted in *Good Thinking* (1983).

GOOD, I. J. (1968c). A five-year plan for automatic chess. In *Machine Intelligence* (E. Dale and D. Michie, eds.) **2** 89–118. Oliver and Boyd, London.

GOOD, I. J. (1972). Random thoughts about randomness. In *Boston Studies in the Philosophy of Science* (D. Reidel, ed.) 117–135. Philosophy of Science Association, East Lansing, Mich. Reprinted in *Good Thinking* (1983).

GOOD, I. J. (1976b). Hallucinations and evidence. *Comm. Statist. B—Simulation Comput.* **B5** 143.

GOOD, I. J. (1979). On the combination of judgments concerning quantiles of a distribution with potential application to the estimation of mineral resources. *J. Statist. Comput. Simulation* **9** 77–79.

GOOD, I. J. (1988d). Yet another argument for the explication of weight of evidence. *J. Statist. Comput. Simulation.* To appear.

GOOD, I. J. (1988e). A suspicious feature of the Popper/Miller argument. *J. Statist. Comput. Simulation.* To appear.

HALDANE, J. B. S. and SMITH, C. A. B. (1947). A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. *Ann. Eugenics* **14** 10–31.

KENDALL, M. G. (1941). A theory of randomness. *Biometrika* **32** 1–5.

LAPLACE, P. S. (1820). *Théorie analytique des probabilités*, 3rd ed. Courcier, Paris.

LINDLEY, D. V., TVERSKY, A. and BROWN, R. V. (1979). On the reconciliation of probability assessments (with discussion). *J. Roy. Statist. Soc. Ser. A* **142** 146–180.

MICHIE, D. (1987). Personal models of rationality. In *Foundations and Philosophy of Probability and Statistics, an International Symposium in Honor of I. J. Good on the Occasion of His 70th Birthday, May 25–26* (K. Hinklemann, ed.). *J. Statist. Plann. Inference.* To appear.

PITMAN, E. J. G. (1965). Some remarks on statistical inference. In *Bernoulli, Bayes, Laplace* (J. Neyman and L. Le Cam, eds.) 215–216. Springer, New York.

SCOTT, C. S. O'D. (1958). A review of a book by G. S. Brown. *J. Soc. Psychic Res.* **39** 217–234.

WRIGHT, S. (1968). *Evolution and the Genetics of Populations. Genetics and Biometric Foundation* **1**. Univ. Chicago Press, Chicago.