7

Dynamic Probability, Computer Chess, and the Measurement of Knowledge*

I. J. Good

Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Philosophers and "pseudognosticians" (the artificial intelligentsia[1]) are coming more and more to recognize that they share common ground and that each can learn from the other. This has been generally recognized for many years as far as symbolic logic is concerned, but less so in relation to the foundations of probability. In this essay I hope to convince the pseudognostician that the philosophy of probability is relevant to his work. One aspect that I could have discussed would have been probabilistic causality (Good, 1961/62), in view of Hans Berliner's forthcoming paper "Inferring causality in tactical analysis", but my topic here will be mainly dynamic probability.

The close relationship between philosophy and pseudognostics is easily understood, for philosophers often try to express as clearly as they can how people make judgments. To parody Wittgenstein, what can be said at all can be said clearly *and it can be programmed.*

A paradox might seem to arise. Formal systems, such as those used in mathematics, logic, and computer programming, can lead to deductions outside the system only when there is an input of assumptions. For example, no probability can be numerically inferred from the axioms of probability unless some probabilities are assumed without using the axioms: *ex nihilo nihil fit.*[2] This leads to the main controversies in the foundations of statistics: the controversies of whether intuitive probability[3] should be used in statistics and, if so, whether it should be logical probability (credibility) or subjective (personal). We who talk about the probabilities of hypotheses, or at least the relative probabilities of pairs of hypotheses (Good, 1950,1975) are obliged to use intuitive probabilities. It is difficult or impossible to lay down precise rules for specifying the numerical values of these probabilities, so some of us emphasize the need for subjectivity, bridled by axioms. At least one of us is convinced, and has repeatedly emphasized for the last thirty years, that a subjective probability can usually be judged only to lie in some interval of values, rather than having a

*Reprinted from *Firbush News*, with the permission of the editor.

sharp numerical value (Good, 1950). This approach arose as a combination of those of Keynes (Keynes, 1921) and of Ramsey (Ramsey, 1931); and Smith's (Smith, 1961) proof of its validity based on certain desiderata, was analogous to the work of Savage (Savage, 1954) who used sharp probabilities.

It is unfortunately necessary once again to express this theory of "comparative subjective probability" in a little more detail before describing the notion of dynamic probability. The theory can be described as a "black box" theory, and the person using the black box is called "you." The black box is a formal system that incorporates the axioms of the subject. Its input consists of your collection of judgments, many of which are of the form that one probability is not less than another one, and the output consists of similar inequalities better called "discernments." The collection of input judgments is your initial *body of beliefs, B*, but the output can be led back into the input, so that the body of beliefs grows larger as time elapses. The purpose of the theory is to enlarge the body of beliefs and to detect inconsistencies in it. It then becomes your responsibility to resolve the inconsistencies by means of more mature judgment. The same black box theory can be used when utilities are introduced and it is then a theory of rationality (Good, 1950,1952).

This theory is not restricted to rationality but is put forward as a model of *all* completed scientific theories.

It will already be understood that the black box theory involves a time element; but, for the sake of simplicity in many applications, the fiction is adopted (implicitly or explicitly) that an entirely consistent body of beliefs has already been attained. In fact one of the most popular derivations of the axioms of probability is based on the assumption that the body of beliefs, including judgments of "utilities" as well as probabilities, is consistent.[4]

One advantage of assuming your body of beliefs to be consistent, in a static sense, is that it enables you to use conventional mathematical logic, but the assumption is not entirely realistic. This can be seen very clearly when the subject matter is mathematics itself. To take a trivial, but very clear example, it would make sense for betting purposes to regard the probability as 0.1 that the millionth digit of π is a 7, yet we know that the "true probability" is either 0 or 1. If the usual axioms of intuitive probability are assumed, together with conventional static logic, it is definitely inconsistent to call the probability 0.1. If we wish to avoid inconsistency we must change the axioms of probability or of logic. Instead of assuming the axiom that $P(E|H) = 1$ when H logically implies E, we must assume that $P(E|H) = 1$ when we have *seen* that H logically implies E. In other words probabilities can change in the light of calculations or of pure thought without any change in the *empirical* data (cf. Good, 1950, p. 49, where the example of chess was briefly mentioned). In the past I have called such probabilities "sliding," or "evolving," but I now prefer the expression *dynamic probability*.[5] It is difficult to see how a subjective probability, whether of a man or of a machine, can be anything other than a dynamic one. We use dynamic probability whenever we make judgments about the truth or falsity of mathematical theorems, and competent mathematicians do this frequently, though usually only informally. There is a naive view that mathematics is concerned only with rigorous logic, a view that arises because finished mathematical proofs *are* more or less rigorous. But in the process of finding and conjecturing theorems every real mathematician is guided by his judgments of what is probably true.[6] This must have been known for centuries, and has been much emphasized and exemplified by Polya (Polya, 1941,1954). A good "heuristic" in problem solving is one that has a reasonable chance of working.[7]

Once the axioms of probability are changed, there is no genuine inconsistency. We don't have to say that P(E|H) has more than one value, for we can denote its value at time t by $P_t(E|H)$, or we can incorporate a notation for the body of beliefs B_t if preferred. There is an analogy with the FORTRAN notation, as in x = x + 3, where the symbol x changes its meaning during the course of the calculation without any real inconsistency.[8]

Believing, as I did (and still do), that a machine will ultimately be able to simulate all the intellectual activities of any man, if the machine is allowed to have the same mechanical help as the man,[9] it used to puzzle me how a machine could make probability judgments. I realized later that this is no more and no less puzzling than the same question posed for a man instead of a machine. We *ought* to be puzzled by how judgments are made, for when we know how they are made we don't call them judgments (Good, 1959B).[10] If judgments ever cease then there will be nothing left for philosophers to do. For philosophical applications of dynamic probability see Appendix A.

Although dynamic probability is implicitly used in most mathematical research it is even more clearly required in the game of chess.[11] For in most chess positions we cannot come as close to certainty as in mathematics. It could even be reasonably argued that the sole purpose of analyzing a chess position, in a game, is for the purpose of improving your estimate of the dynamic probabilities of winning, drawing, or losing. If analysis were free, it would pay you in expectation to go on analyzing until you were blue in the face, for it is known that free evidence is always of non-negative expected utility (for example, (Good, 1967A), but see also (Good, 1974)). But of course analysis is not free, for it costs effort, time on your chess clock, and possibly facial blueness. In deciding formally how much analysis to do, these costs will need to be quantified.

In the theory of games, as pioneered mainly by von Neumann (von Neumann, 1944/47), chess is described as a "game of perfect information," meaning that the rules involve no reference to dice and the like. But in practice most chess positions cannot be exhaustively analyzed by any human or any machine, *present or future.*[12] Therefore play must depend on probability even if the dependence is only implicit. Caissa is a cousin of the Moirai after all.

Against this it can be argued that the early proposals for helping humans and computers to play chess made use of evaluation functions (for quiescent positions) and did not rely on probability, dynamic or otherwise. For example, the beginner is told the value of the pieces, P = 1, B = 3.25, etc. and that central squares are usually more valuable than the others.[13] But an evaluation function can be fruitfully interpreted in probabilistic terms and we now recall a conjectured approximate relationship that has been proposed (Good, 1959B,1967B) by analogy with the technical definition of weight of evidence.

The weight of evidence, provided by observations E, in favour of one hypothesis H_1, as compared with another one H_2, is defined as

$$\log \frac{O(H_1/H_2|E)}{O(H_1/H_2)} = \log \frac{P(E|H_1)}{P(E|H_2)}$$

where P denotes probability and O denotes odds. In words, the weight of evidence, when added to the intial log-odds, gives the final log-odds. The expression "weight of evidence," in this sense, was used independently in (Pierce, 1878), (Good, 1950), and (Minsky and Selfridge, 1961). Weight of evidence has simple additive and other properties which make it, in my opinion, by far the best explicatum for corroboration (Good, 1960/68, 1968, 1975). The conjecture is that *ceteris paribus* the weight of evidence in favour of White's winning as compared with losing, in a given position, is roughly proportional to her advantage in material, or more generally to the value of her evaluation function, where the constant of proportionality will be larger for strong players than for weak ones. The initial log-odds should be defined in terms of the playing strengths of the antagonists, and on whether the position is far into the opening, middle-game, or end-game, etc. Of course this conjecture is susceptible to experimental verification or refutation or improvement by statistical means, though not easily; and at the same time the conjecture gives additional meaning to an evaluation function.[14] As an example, if an advantage of a pawn triples your odds of winning as compared with losing, then an advantage of a bishop should multiply your odds by about $3^{3.25} = 35.5$. This quantitative use of probability is not in the spirit of Polya's writings, even if interval estimates of the probabilities are used.

If dynamic probability is to be used with complete seriousness, then it must be combined with the principle of rationality (see Appendix A). First you should decide what your utilities are for winning, drawing, and losing, say u_W, u_D, and u_L. More precisely, you do not need all three parameters, but only the ratio $(u_W - u_D)/(u_D - u_L)$. Then you should aim to make the move, or one of the moves, that maximize the mathematical expectation of your utility, in other words you should aim to maximze

$$p_W u_W + p_D u_D + p_L u_L \tag{1}$$

where p_W, p_D, and p_L are your dynamic probabilities of winning, drawing, or losing. When estimating (1) you have to allow for the state of the chess clock so that the "costs of calculation," mentioned in Appendix A, are very much in the mind of the match chess player.[15] This is not quite the whole picture because you might wish to preserve your energy for another game: this accounts for many "grandmaster draws."

Current chess programs all depend on tree analysis, with backtracking, and the truncation of the tree at certain positions. As emphasized in (Good, 1967B), it will eventually be necessary for programs to handle descriptions of positions[16] if Grandmaster status is to be achieved, and the lessons derived from this work will of course change the world, but we do not treat this difficult matter in this paper.

For the moment let us suppose that the problem has been solved of choosing the nodes where the tree is to be truncated. At each such node the probabilities p_W, p_D, and p_L are a special kind of dynamic probability, namely *superficial* or *surface* probabilities, in the sense that they do not depend on an analysis in depth. The evaluation function used at the end-nodes, which is used for computing these three probabilities, might depend on much deep cogitation and statistical analysis, but this is not what is meant here by an "analysis in depth." Then the minimax backtracking procedure can be used; or expectimaxing if you wish to allow for the deficiencies of your opponent,[17] and for your own deficiencies. In this way you can arrive at values of the dynamic probabilities $p_W{}^0$, $p_D{}^0$, and $p_L{}^0$ corresponding to the positions that would arise after each of your plausible moves in the *current* position, π_0. Of course these probabilities depend on the truncation rules (pruning or pollarding).

Some programs truncate the analysis tree at a fixed depth but this is very unsatisfactory because such programs can never carry out a deep combination. Recognizing this, the earliest writers on chess programming, as well as those who discussed chess programming intelligently at least ten years earlier,[18] recognized that an important criterion for a chess position π to be regarded as an endpoint of an analysis tree was *quiescence*. A quiescent position can be defined as one where the player with the move is neither threatened with immediate loss, nor can threaten his opponent with immediate loss. The primary definition of "loss" here is in material terms, but other criteria should be introduced. For example, the advance of a passed pawn will often affect the evaluation non-negligibly. We can try to "materialize" this effect, for example, by regarding the value of a passed pawn, not easily stopped, as variable. My own proposals are 1¼ on the fourth rank, 1½ on the fifth rank, 3 on the sixth, 5 on the seventh, and 9 on the eighth!, but this is somewhat crude.[19]

An informal definition of a turbulent position is a combinative one. For example, the position: White K at c3, R at c8; Black K at al, R at a4; is turbulent. But if Black also has a Q at f6, then White's game is hopeless, so the turbulence of the position does not make it much worth analyzing.

Hence in (Good, 1967B, p. 114) I introduced a term *agitation* to cover both turbulence and whether one of p_W, p_D, and p_L is close to 1. Apart from considering whether to threaten to take a piece, in some potential future position π, we should consider whether the win of this piece would matter much. Also, instead of considering one-move threats, it seems better to consider an analysis of unit cost, which might involve several moves, as, for example, when checking whether a pawn can be tackled before it touches down in an end-game. The definition of the *agitation* $A(\pi)$ was the expected value of $|U(\pi|\$) - U(\pi)|$ where $U(\pi)$ is the superficial utility of π and $U(\pi|\$)$ is the utility of π were a unit of amount of analysis to be done. ($U(\pi|\$)$ is a subjective random variable before the analysis is done.)

But the depth from the present position π_0 to π is also relevant in the decision whether to truncate at π. More exactly, the dynamic probability $P(\pi|\pi_0)$ that position π will be reached from π_0 is more relevant than the depth. We could even reasonably define the probabilistic depth as proportional to $-\log P(\pi|\pi_0)$ and the effective depth of the whole analysis as $-\Sigma P(\pi|\pi_0) \log P(\pi|\pi_0)$, summed over all endpoints π, as suggested in (Good, 1967B). But the most natural criterion for whether to treat π as an endpoint in the analysis of π_0 is obtained by setting a threshold on $P(\pi|\pi_0)A(\pi)$. The discussion of agitation and allied matters is taken somewhat further in (Good, 1967B, pp. 114-115).

As a little exercise on dynamic probability let us consider the law of multiplication of advantage which states that "with best play on both sides we would expect the rate of increase of advantage to be some increasing function of the advantage." This might appear to contradict the conjecture that the values of the pieces are approximately proportional to weights of evidence in favour of winning rather than losing. For we must have the "Martingale property" $E(p_t|p_0) = p_0$, where p_0 and p_t are the probabilities of winning at times 0 and t. This only sounds paradoxical if we forget the elementary fact that the expectation of a function is not usually equal to that same function of the expectation. For example, we could have, for some $\epsilon > 0$,

$$E(\log \frac{p_t}{1-p_t}) \approx (1 + \epsilon)^t \log \frac{p_0}{1-p_0} \qquad (2)$$

without contradicting the Martingale property, and (2) expresses a possible form of the law of multiplication of advantage, though it cannot be very accurate.

An idea closely associated with the way that dynamic probabilities can vary is the following method for trying to improve any given chess program. Let the program starting in a position π_0, play against itself, say for the next n moves, and then to quiescence, at say π_1. Then the odds of winning from position π_1, or the expected utility, could be used for deciding whether the plan and the move adopted in position π_0 turned out well or badly. This information could be used sometimes to change the decision, for example, to eliminate the move chosen before revising the analysis of π_0. This is not the same as a tree analysis alone, starting from π_0, because the tree analysis will often not reach the position π_1. Rather, it is a kind of learning by experience. In this procedure n should not be at all large because non-optimal moves would introduce more noise the larger n was taken. The better the standard of play the larger n could be taken. If the program contained random choices, the decision at π_0 could be made to depend on a *collection* of sub-games instead of just one. This idea is essentially what humans use when they claim that some opening line "appears good in master practice."

To conclude this paper I should like to indicate the relevance of dynamic probability to the quantification of knowledge, for which Michie proposed a non-probabilistic measure.[20] As he points out, to know that $12^3 = 1728$ can be better than having to calculate it, better in the sense that it saves time. His discussion was non-probabilistic so it could be said to depend, at least implicitly, on dynamic *logic* rather than on dynamic *probability*. In terms of dynamic probability, we could describe the *knowledge* that $12^3 = 1728$ as the ascribing of dynamic probability $p = 1$ to this mundane fact. If instead p were less than 1, then the remaining dynamic information available by calculation would be $-\log p$ (Good, 1950, p. 75; Good, 1968, p. 126). This may be compared with Michie's definition of amount of knowledge, which is based on Hartley's non-probabilistic measure of information (Hartley, 1928).

Amount of knowledge can be regarded as another quasi-utility of which weight of evidence and explicativity are examples. A measure of knowledge should be usable for putting programs in order of merit.

In a tree search, such as in chess, in theorem-proving, and in medical diagnosis, one can use entropy, or amount of information, as a quasi-utility for cutting down on the search (Good, 1970; Good and Card, 1971; Card and Good, 1974) and the test for whether this quasi-utility is sensible is whether its use agrees reasonably well with that of the principle of rationality, the maximization of expected utility. Similarly, to judge whether a measure of knowledge is a useful quasi-utility it should ultimately by compared with the type 2 principle of rationality (Appendix A). So the question arises what form this principle would take when applied to computer programs.

Suppose we have a program for evaluating a function f(x) and let's imagine for the moment that we are going to make one use of the program for calculating f(x) for some unknown value of x. Suppose that the probability that x will be the value for which we wish to evaluate the function is p(x) and let's suppose that when we wish to do this evaluation the utility of the calculation is $u(x,\lambda)$ where λ is the proportional accuracy of the result. Suppose further that the cost of obtaining this proportional accuracy for evaluating f(x), given the program, is $c(x,\lambda)$. Then the total expected utility of the program, as far as its next use is concerned, is given by the expression

$$\left.\begin{aligned} U &= \int p(x) \max\left\{0, \max_{\lambda}[u(x,\lambda) - c(x,\lambda)]\right\} dx \\ \text{or}\quad & \\ &\sum_{x} p(x) \max\left\{0, \max_{\lambda}[u(x,\lambda) - c(x,\lambda)]\right\}. \end{aligned}\right\} \tag{3}$$

The notion of dynamic probability (or of rationality of type 2) is implicit in the utilities mentioned here, because, if the usual axioms of probability are assumed, the utilities would be zero because the costs of calculation are ignored. Anything calculable is "certain" in ordinary logic and so conveys no logical information, only dynamic information.

If the program is to be applied more than once, then formula (3) will apply to each of its applications unless the program is an adaptive one. By an *adaptive program* we could mean simply that *the costs of calculation tend to decrease* *when the program is used repeatedly.* This will be true for example, in the adaptive rote learning programs that Donald Michie described in lectures in Blacksburg in 1974. To allow for adaptability would lead to severe complications and I suspect that similar complications would arise if Donald Michie's definition of amount of knowledge were to be applied to the same problem.

I expect his definition will usually be considerably easier to use then the expression (3), but I do not know which is the better definition on balance.

Example 1. Suppose that (i) all accurate answers have a constant utility a, and all others have zero utility. Then

$$u(x,\lambda) = \begin{cases} a & \text{if } \lambda = 0 \\ 0 & \text{otherwise;} \end{cases}$$

(ii) $c(x,0) = b$, a constant, when x belong to a set X, where $a > b > 0$, and that $c(x,0) > a$ if x does not belong to X; (iii) all values of x are equally likely a priori, that is, $p(x)$ is mathematically independent of x. Then (3) is proportional to the number of elements in X, that is, to the number of values of x that can be "profitably" computed.

Example 2.

$$u(x,\lambda) = \begin{cases} a & \text{if } \lambda < \lambda_0 \\ 0 & \text{otherwise.} \end{cases}$$

The analysis is much the same as for Example 1 and is left to the reader.

Example 3. $u(x,\lambda) = -\log \lambda (\lambda < 1)$; then the utility is approximately proportional to the number of correct significant figures.

Example 4. In the theory of numbers we would often need to modify the theory and perhaps use a utility $u(x,\mu)$, where μ is the number of decimal digits in the answer.

Example 5: knowledge measurement in chess. Let x now denote a chess position instead of a number. Let $u(x)$ denote the expected utility of the program when applied in position x, allowing this time for the costs. Then $v = \Sigma_x p(x)u(x)$ measures the expected utility of the program per move, where $p(x)$ is the probability of the occurrence of position x. The dependence between consecutive positions does not affect this formula because the expectation of a sum is always the sum of the expectations regardless of dependence. A measure of the knowledge added to the program by throwing the book on opening variations at it, can be obtained by simply subtracting the previous value of v from the new value.

It should now be clear that dynamic probability is fundamental for a theory of practical chess, and has wider applicability. Any search procedure, such as is definitely required in non-routine mathematical research, whether by humans or by machines, *must* make use of subgoals to fight the combinatorial explosion. Dynamic utilities are required in such work because, *when you set up subgoals, you should estimate their expected utility as an aid to the main goal* before you bother your pretty head in trying to attain the subgoals.

The combinatorial explosion is often mentioned as a reason for believing in the impracticability of machine intelligence, but if this argument held water it would also show that human intelligence is impossible. Perhaps it is impossible for a human to be intelligent, but the real question is whether machines are necessarily equally unintelligent. Both human problem-solvers and pseudo-gnostical machines must use dynamic probability.

Appendix A. Philosophical applications of dynamic probability

An interesting application of dynamic probability is to a fundamental philosophical problem concerning simplicity. Many of us believe that of two scientific laws that explain the same facts, the simpler is usually the more probable. Agassi, in support of a thesis of Popper, challenged this belief by pointing out that, for example, Maxwell's equations imply Fresnel's optical laws and must therefore be not more probable, yet Maxwell's equations appear simpler. This difficulty can be succinctly resolved in terms of dynamic probability, and I believe this is the only possible way of resolving it. For the clarification of these cryptic remarks see (Good, 1968) and (Good, 1975). These papers also contain an explication and even a calculus for "explicativity," a quantitative measure of the explanatory power of a theory.

A further philosophical application of dynamic probability arises in connection with the principle of rationality, the recommendation to maximize expected utility. It frequently happens that that amount of thinking or calculation required to obey this principle completely is very great or impracticably large. Whatever its size, it is rational to allow for the costs of this effort (for example, [Good, 1971]), whatever the difficulties of laying down rules for doing so. When such allowance is made we can still try to maximize expected utility, but the probabilities, and sometimes the utilities also, are then dynamic. When a conscious attempt is made to allow for the costs we may say we are obeying the principle of rationality of type 2. This modified principle can often justify us in using the often convenient but apparently *ad hoc* and somewhat irrational methods of "non-Bayesian" statistics, that is, methods that officially disregard the use of subjective probability judgments. But such judgments are always at least implicit: all statisticians are implicit Bayesians whether they know it or not, except sometimes when they are making mistakes. (Of course Bayesians also sometimes make mistakes.)

Thus dynamic probability and dynamic utility help us to achieve a Bayes/non-Bayes synthesis. Inequality judgments rather than sharp probability judgments also contribute to this synthesis: a strict non-Bayesian should choose the interval (0,1) for all his subjective probabilities! For an interesting example of a Bayes/non-Bayes synthesis see (Good, 1967C) and (Good and Crook, 1974).

NOTES

1. Lighthill's joke, cracked in a BBC TV debate. Jokes don't wear well for long, however

risible they were originally, so I have invested a neologism that just might replace the clumsy and ambiguous "workers in A.I." The "g" of "pseudognostics" belongs to the third syllable! Michie's expression "knowledge engineering" might be preferred in some contexts, but it will tend to prevent A.I. work in any university department outside engineering. Engineering departments already tend to take the universities over.

2. Each axiom merely *relates* probability values. *Suggestions*, such as the "principle of sufficient reason," are not axioms and they require judgments about the real world.
3. By "intuitive probability" I mean either logical or subjective probability (Koopman, 1940) as contrasted with the physical probabilities that arise, for example, in quantum mechanics, or the tautological probabilities of mathematical statistics (Good, 1959A).
4. More precisely, it must be "coherent" in the sense that a "Dutch book" cannot be made against it in a gambling situation. A Dutch book is a proposed set of bets such that you will lose whatever happens (Savage, 1954).
5. Donald Michie expressed a preference for this term in conversation in 1974, since he thought that "evolving probability," which I have used in the past, was more likely to be misunderstood.
6. (i) A real mathematician, by definition, cannot do all his work by low-level routine methods; but one man's routine is another man's creativity. (ii) Two famous examples of the use of *scientific* induction in mathematics were Gauss's discoveries of the prime number theorem and of the law of quadratic reciprocity. He never succeeded in proving the first of these results.
7. Polya's writings demonstrate the truth of the aphorism in Note 6. Polya's use of probability in mathematical research is more qualitative than mine. A typical theorem in his writings is "The more confidence we placed in an incompatible rival of our conjecture, the greater will be the gain of faith in our conjecture when that rival is refuted" (Polya, 1954, vol 2, p. 124). His purely qualitative approach would prevent the application of the principle of rationality in many circumstances.
8. Presumably the ALGOL notation x: = x + 3 was introduced to avoid the apparent inconsistency.
9. It is pointless to make such judgments without some attached dynamic probabilities, so I add that I think there is a probability exceeding ½ that the machine will come in the present century. But a probability of only 1/1000 would of course justify the present expenditures.
10. Judgments are never formalized
You can sign that with your blood gents
For when they are formalized
No one dare call them judgments.

Drol Doog (With apologies to Sir John Harrington.)

11. In case this seems too obvious the reader is reminded that it was not explicit in the earlier papers on chess programming, and there is no heading "Probability" in (Sunnucks, 1970).
12. Even if every atom in the moon examined 10^{24} *games* per second (light takes about 10^{-24} sec. to traverse the diameter of an electron), it would take ten million times the age of the universe to examine 10^{100} games, which is a drop in the mare.
13. The values of the pieces also vary with the position, in anyone's book. There is much scope for conjectures and statistical work on evaluation functions. For example, it was suggested in (Good, 1967B) that the "advantage of two bishops" could be explained by assuming that it is "in general" better to control two different squares than to control one square twice, although "overprotection of the centre" might be an exception. For example, the contribution to the total "score" from the control n times of one square might be roughly proportional to $(n + 1)^{\alpha} - (n + 1)^{-\beta}$ $(0 < \alpha < 1, \beta > 0)$.
14. Perhaps the odds of a draw are roughly the geometric mean of those of winning and losing.

15. The International Chessmaster and senior Civil Servant, Hugh Alexander, once remarked that it is more important for a Civil Service administrator to make his mind up promptly than to reach the best decision. He might have had in mind that otherwise the administrator would "lose on the clock."
16. To be precise I said that natural language should be used, and John McCarthy said from the floor that descriptions in symbolic logic might be better.
17. This is known as psychological chess when Emanual Lasker does it, and trappy chess when I do it.
18. By definition of "intelligently."
19. The difficulty of evaluating unblocked passed pawns is one for the human as well as for the machine, because it is often in the balance whether such pawns *can* be blocked. This might be the main reason for the difficulty of formalizing endgame play. It is said that mathematicians have an advantage in the endgame but I do not know the evidence for this nor clearly why it should be true.
20. This part of the paper is based on my invited discussion of Michie's public lecture on the measurement of knowledge on October 30, 1974 in Blacksburg: see his contribution to this volume.

REFERENCES

Card, W.I. and I.J. Good (1974) A logical analysis of medicine. *A Companion to Medical Studies* (eds. R. Passmore and J.S. Robson; Oxford, Blackwell's), 3, Chapter 60.

Good, I.J. (1950) *Probability and the Weighing of Evidence* (London, Charles Griffin; New York, Hafners), p. 119.

Good, I.J. (1952) Rational decisions. *J. Roy. Statist. Soc. B*, **14**, 107-114.

Good, I.J. (1959A) Kinds of probability. *Science* **129**, 443-447.

Good, I.J. (1959B) Could a machine make probability judgments? *Computers and Automation* 8, 14-16 and 24-27. (Based on a lecture at the British Society for the Philosophy of Science, Jan. 1958.)

Good, I.J. (1960/68) Weight of evidence, corroboration, explanatory power, information, and the utility of experiments. *J. Roy. Statist. Soc.-B*, **22**, 319-331; 30, 203.

Good, I.J. (1961/62) A causal calculus *Brit. J. Philos. Sc.* **11**, 305-318; **12**, 43-51; **13**, 88.

Good, I.J. (1967A) On the principle of total evidence. *Brit J. Philos. Sc.* **17**, 319-321.

Good, I.J. (1967B) A five-year plan for automatic chess. *Machine Intelligence* **2** (eds. E. Dale and D. Michie; Oliver and Boyd), 89-118.

Good, I.J. (1967C) A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. B.*, **29**,339-431 (with discussion).

Good, I.J. (1968) Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *Brit. J. Philos. Sc.* **19**, 123-143. See also **25** (1974), 289.

Good, I.J. (1970) Some statistical methods in machine intelligence research. *Math. Biosc.* **6**, 185-208. (Based on a lecture at Fort Belvoir, November 1967.)

Good, I.J. (1971) Twenty-seven principles of rationality, (Appendix to "The probabilistic explication of information, evidence, surprise, causality, explanation, and utility"). *Foundations of Statistical Inference* (eds. V.P. Godambe and D.A. Sprott; Holt, Rinehart and Winston of Canada, Ltd., Toronto and Montreal), 124-127.

Good, I.J. (1974) A little learning can be dangerous. *Brit J. Philos. Sc.* **25**, 340-342.

Good, I.J. (1975) Explicativity, corroboration, and the relative odds of hypotheses. *Synthese*, **30**, 39-73.

Good, I.J. and W.I. Card (1971) The diagnostic process with special reference to errors. *Methods of Information in Medicine* **10**, No. 3, 176-188.

Good, I.J. and J.F. Crook (1974) The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Soc.* **69**, 711-720.

Hartley, R.V.L. (1928) Transmission of information. *Bell System Tech. J.* **7**, 535-563.

Horowitz, I.A. and G. Mott-Smith (1960) *Point Count Chess* (New York: Simon and Schuster).

Keynes, J.M. (1921) *A Treatise on Probability* (London: Macmillan).

Koopman, B.O. (1940) The basis of probability. *Bull. Amer. Math. Soc.* 46, 763-774; The axioms and algebra of intuitive probability. *Annals of Math.* 41, 269-292.

Minsky, M. and O.G. Selfridge (1961) Learning in random nets, in *Information Theory* (ed. Colin Cherry; London: Butterworths), 335-347.

von Neumann, J. and O. Morganstern (1944/47) *The Theory of Games and Economic Bahavior* (Princeton: University Press).

Pierce, C.S. (1878) The probability of induction. *Popular Science Monthly*, reprinted in *The World of Mathematics*, 2 (ed. James R. Newman; New York: Simon and Schuster, 1956), 1341-1354.

Polya, G. (1941) Heuristic reasoning and the theory of probability. *Amer. Math. Monthly*, 48, 450-465.

Polya, G. (1954) *Mathematics and Plausible Reasoning*, two vols. (Princeton: University Press).

Ramsey, F.P. (1931) *The Foundations of Mathematics* (London: Kegan Paul).

Savage, L.J. (1954) *The Foundation of Statistics* (New York and London: Wiley).

Smith, C.A.B. (1961) Consistency in statistical inference and decision, *J. Roy. Statist. Soc. B*, 23, 1-25.

Sunnucks, Anne (1970) *The Encyclopaedia of Chess* (New York: St. Martin's Press).