Twenty-seven Principles of Rationality (#679)

In the body of my paper for this symposium I originally decided not to argue the case for the use of subjective probability since I have expressed my philosophy of probability, statistics, and (generally) rationality on so many occasions in the past. But after reading the other papers I see that *enlightenment* is still required. So, in this appendix I give a succinct list of 27 priggish principles. I have said and stressed nearly all of them before, many in my 1950 book, but have not brought so many of them together in one short list. As Laplace might have said, they are *au fond le bon sens*, but they cannot be entirely reduced to a calculus. [The writers who influenced me are mentioned in the Introduction.]

1. Physical probabilities probably *exist* (I differ from de Finetti and L. J. Savage here) but they can be *measured* only with the help of subjective probabilities. There are several kinds of probability. (##**182**, 522. The latter paper contains a dendroidal categorization.)

2. A familiar set of axioms of subjective probability are to be used. Kolmogorov's axiom (complete additivity) is convenient rather than essential. The axioms should be combined with rules of application and less formal suggestions for aiding the judgment. Some of the suggestions depend on theorems such as the laws of large numbers which make a frequency definition of probability unnecessary. (It is unnecessary and insufficient.)

3. In principle these axioms should be used in conjunction with inequality judgments and therefore they often lead only to inequality discernments. The axioms can themselves be formulated as inequalities but it is easier to incorporate the inequalities in the rules of application. In other words most subjective probabilities are regarded as belonging only to some interval of values the end points of which may be called the lower and upper probabilities. (Keynes, 1921; Koopman, 1940a; ##13, **230**; C. A. B. Smith, 1961.)

4. The principle of rationality is the recommendation to maximize expected utility.

5. The input and output to the abstract theories of probability and rationality are judgments of inequalities of probabilities, odds, Bayesian factors (ratios of final to initial odds), log-factors or weights of evidence, information, surprise indices, utilities, *and any other functions of probabilities and utilities.* (For example, Good, 1954.) It is often convenient to forget about the inequalities for the sake of simplicity and to use precise estimates (see Principle 6).

6. When the expected time and effort taken to think and do calculations are allowed for in the costs, then one is using the principle of *rationality of type II.* This is more important than the ordinary principle of rationality, but is seldom mentioned because it contains a veiled threat to conventional logic by incorporating a time element. It often justifies *ad hoc* [and non-Bayesian] procedures such as confidence methods and this helps to decrease controversy.

7. The purposes of the theories of probability and rationality are to enlarge bodies of beliefs and to check them for consistency, and thus to improve the objectivity of subjective judgments. This process can never be [the words "known to be" should be inserted here to cope with Dr. Barnard's comment that followed this paper] completed even in principle, in virtue of Gödel's theorem concerning consistency. Hence the type II principle of rationality is a logical necessity.

8. For clarity in your own thinking, and especially for purposes of communication, it is important to state what judgments you have used and which parts of your argument depend on which judgments. The advantage of likelihood is its mathematical independence of initial distributions (priors), and similarly the advantage of weight of evidence is its mathematical independence of the initial odds of the null hypothesis. The subjectivist states his judgments whereas the objectivist sweeps them under the carpet by calling assumptions *knowledge*, and he basks in the glorious objectivity of science.

9. The vagueness of a probability judgment is defined either as the difference between the upper and lower probabilities or else as the difference between the upper and lower log-odds (#290). I conjecture that the vagueness of a judgment is strongly correlated with its variation from one judge to another. This could be tested.

10. The distinction between type I and type II rationality is very similar to the distinction between the standard form of subjective probabilities and what I call *evolving* or *sliding* [or *dynamic*] *probabilities.* These are probabilities that are currently judged and they can change in the light of thinking only, without change of empirical evidence. The fact that probabilities change when *empirical* evidence changes is almost too elementary a point to be worth mentioning in this distinguished assembly, although it was overlooked by . . . R. A. Fisher in his fiducial argument. More precisely he talked about the probabilities of certain events or propositions without using the ordinary notation of the vertical stroke or any corresponding notation, and thus fell into a fallacy. . . . Evolving probabilities are essential for the refutation of Popper's views on simplicity (see #599 [or p. 223]). [Great men are not divine.]

11. My theories of probability and rationality are theories of consistency

only, that is, consistency between judgments and the basic axioms and rules of application of the axioms. Of course, these are usually judgments about the objective world. In particular it is incorrect to suppose that it is necessary to inject an initial probability distribution from which you are to infer a final probability distribution. It is just as legitimate logically to assume a final distribution and to infer from it an initial distribution. (For example, pp. 35 and 81 of #13.) To modify an aphorism quoted by Dr. Geisser elsewhere, "Ye priors shall be known by their posteriors."

12. This brings me to the *device of imaginary results* (#13) which is the recommendation that you can derive information about an initial distribution by an imaginary (*Gedanken*) experiment. Then you can make discernments about the final distribution after a real experiment. . . .

13. *The Bayes/non-Bayes compromise* (p. 863 of #127, and also many more references in #398 under the index entry "Compromises"). Briefly: use Bayesian methods to produce statistics, then look at their tail-area probabilities and try to relate these to Bayes factors. A good example of both the device of imaginary results and of the Bayes/non-Bayes compromise was given in #547. I there found that Bayesian significance tests in multiparameter situations seem to be much more sensitive to the assumed initial distribution than Bayesian estimation is.

14. The weakness of Bayesian methods for significance testing is also a strength, since by trying out your assumed initial distribution on problems of significance testing, you can derive much better initial distributions and these can then be used for problems of estimation. This improves the Bayesian methods of estimation!

15. Compromises between subjective probabilities and credibilities are also desirable because standard priors might be more general-purpose than non-standard ones. In fact it is mentally healthy to think of your subjective probabilities as estimates of credibilities (p. 5). Credibilities are an ideal that we cannot reach.

16. The need to compromise between simplicity of hypotheses and the degree to which they explain the facts was discussed in some detail in #599, and the name I gave for the appropriate and formally precise compromise was "a Sharpened Razor." Ockham (actually his eminent predecessor John Duns Scotus) in effect emphasized simplicity alone, without reference to *degrees* of explaining the facts. (See also #**1000**.)

17. The relative probabilities of two hypotheses are more relevant to science than the probabilities of hypotheses *tout court* (pp. 60, 83-84 of #13).

18. The objectivist or his customer reaches precise results by throwing away evidence; for example (a) when he keeps his eyes averted from the precise choice of random numbers by using a Statistician's Stooge; (b) when his customer uses confidence intervals for betting purposes, which is legitimate *provided that he regards the confidence statement as the entire summary of the evidence.*

19. If the objectivist is prepared to bet, then we can work backwards to infer

constraints on his implicit prior beliefs. These constraints are of course usually vague, but we might use precise values in accordance with type II rationality.

20. When you don't trust your estimate of the initial probability of a hypothesis you can still use the Bayes factor or a tail-area probability to help you decide whether to do more experimenting (p. 70 of #13).

21. Many statistical techniques are legitimate and useful but we should not knowingly be inconsistent. The Bayesian flavor vanishes when a probability is judged merely to lie in the interval (0,1), but this hardly ever happens.

22. A hierarchy of probability distributions, corresponding in a physical model to populations, superpopulations, etc., can be helpful to the judgment even when these superpopulations are not physical. I call these *distributions of types I, II, III, . . ."* (partly in order to be noncommittal about whether they are physical) but it is seldom necessary to go beyond the third type (##**26**, 398, 547). [When we are prepared to be committal, the current names for types II and III are "priors" and "hyperpriors."]

23. Many compromises are possible, for example, one might use the generalizations of the likelihood ratio mentioned on p. 80 of #198.

24. *Quasi- or pseudoutilities*. When your judgments of utilities are otherwise too wide it can be useful [in the planning of an experiment] to try to maximize the expectation of something else that is of value, known as a quasiutility or pseudoutility. Examples are (a) weight of evidence, when trying to discriminate between two hypotheses; (b) information in Fisher's sense when estimating parameters; (c) information in Shannon's sense when searching among a set of hypotheses; (d) strong explanatory power [explicativity] when explanation is the main aim: this includes example (c); (e) and (ea) tendency to cause (or a measure of its error) if effectiveness of treatment (or its measurement) is the aim; (f) f(error) in estimation problems, where $f(x)$ depends on the application and might, for example, reasonably be $1 - e^{-\lambda x^2}$ (or might be taken as x^2 for simplicity) [the sign should be changed]; (g) financial profit when other aims are too intangible. In any case the costs in money and effort have to be allowed for. [*After* an experiment is designed, a minimax inference minimizes the quasiutility. Minimax procedures cannot be fully justified but they lead to interesting proposals. See #622 or p. 198 of #618.]

25. The time to make a decision is largely determined by urgency and by the current rate of acquisition of information, evolving or otherwise. For example, consider chess timed by a clock.

26. In logic, the probability of a hypothesis does not depend on whether it was typed accidentally by a monkey, or whether an experimenter pretends he has a train to catch when he stops a sequential experiment. But in practice we do allow for the degree of respect we have for the ability and knowledge of the person who propounds a hypothesis.

27. All scientific hypotheses are numerological but some are more numerologi-

cal than others. Hence a subjectivistic analysis of numerological laws is relevant to the philosophy of induction (#603B).

I have not gone systematically through my writings to make sure that the above list is complete. In fact there are, for example, a few more principles listed in ##85A, 293. But I believe the present list is a useful summary of my position.