quency or just falls out of another type of theory, I think we do argue that way. I suspect that even the most extreme subjectivist such as de Finetti would have to agree that he did sometimes think that way, though he would perhaps avoid doing it in print. I do not think there is all that much distinction between the metaphysical status of subjective and physical probabilities. You can arrive at the numerical value of a physical probability by means of a repeated experiment in which you gradually modify the subjective probability and in that way you can measure your physical probability in terms of subjective probabilities.

Savage has shown that a rational man behaves as if he used subjective probabilities. A rational man will also presumably behave as if he thought the world behaves as if there are physical probabilities. When he measures these physical probabilities he will behave as if they were limiting values of his subjective probabilities. Thus both types of probability are metaphysical, and perhaps everything is. I mean we use language and behave *as if* we had various opinions.

Mr D. V. LINDLEY: I think we ought to look carefully at the situations that the subjectivist seems to analyse differently from the way that most of us have been taught to use. One of these situations, described by Professor Savage in Part I, concerns optional stopping. He gave a very pertinent discussion of what happens when we have six successes out of a hundred. I am disappointed that none of the other speakers has been tempted to reply to this, to say whether he would agree with Professor Savage or not. Is there for instance someone who feels that he wants to use estimates that take account of the stopping rule?

BARTLETT: I am not going to answer the question completely, but there is one small point I should like to make. Certainly I agree that unbiased estimates are unimportant. And in the particular problem, I have pointed out before if you have six successes out of a hundred in an ordinary fixed-sample-size situation, you take 6/100 as the sufficient statistic that happens to be unbiased and carries the maximum information. If you have inverse sampling, it seems to me that certainly you should take 100/6 as your unbiased estimate of $1/p$, which is sufficient and carries the maximum information in the inverse sampling case. You have to take 5/99 to get an unbiased estimate of p. I should like to think about the general question further though.

Dr P. ARMITAGE: I think it is quite clear that likelihood ratios, and therefore posterior probabilities, do not depend on a stopping rule. Professor Savage, Dr Cox and Mr Lindley take this necessarily as a point in favour of the use of Bayesian methods. My own feeling goes the other way. I feel that if a man deliberately stopped an investigation when he had departed sufficiently far from his particular hypothesis, then 'Thou shalt be misled if thou dost not know that'. If so, prior probability methods seem to appear in a less attractive light than frequency methods, where one can take into account the method of sampling. I should like Professor Savage to clarify a point he made in Part I. He remarked that, using conventional significance tests, if you go on long enough you can be sure of achieving any level of significance; does not the same sort of result happen with Bayesian methods? The departure of the mean by two standard errors corresponds to the ordinary five per cent level. It also corresponds to the null hypothesis being at the five per cent point of the posterior distribution. Does it not follow that by going on sufficiently long one can be sure of getting the null value arbitrarily far into the tail of the posterior distribution?

SAVAGE: The answer is surely no, under any interpretation. It is impossible to be sure of sampling until the data justifies an unjustifiable conclusion, just as surely as it is impossible to build a perpetual-motion machine. After all, whatever we may disagree about, we are surely agreed that Bayes's theorem is true where it applies. But to understand this impossibility let us examine first a simple case.

Consider an urn that contains three red balls and a black one or three black balls and a red one. To convince you of the first hypothesis as opposed to the second, for some given purpose, would mean to make the likelihood ratio in favour of the first sufficiently large, say at least 10. Suppose that I, in my zeal, decide to keep sampling (with replacement) until the likelihood ratio, which in this particular case is 3^{r-b}, exceeds 10. This will happen if and only if I sometimes succeed in drawing three more red balls than black ones; if there are really three black balls and a red one, it is quite probable that I never will succeed until the end of time. In fact, the probability of failure in this unfavourable circumstance is at least 9/10, as it ought to be on general principles; the exact value is 26/27.

As I understand it, Dr Armitage is particularly interested in the following sort of example. The prior distribution of a parameter μ is rather broadly distributed around 0, and observations of μ with unit standard deviation are sequentially available. From 'your' point of view, that is, the point of view summarized by the assumed prior distribution, what is the probability P that I should succeed in sampling until your posterior odds that μ is positive are at least 10 times your initial odds that μ is positive, if μ is in fact negative? There can be no escape from the simple general formula that P is at most a tenth, but there might be some momentary misunderstanding of the meaning of that formula.

If μ is not negative, and I sample with a determination to raise your odds in favour of the proposition that μ is positive by large factor, I am of course sure to succeed. Still more, if μ is only very slightly negative, then, with determination I am almost sure to succeed in convincing you that μ is positive. This may at first seem objectionable, but you must not forget that 'you' felt very sure at the outset that μ was not close to 0, so the general conclusion that you are not unduly likely to be fooled has not been upset.

If optional stopping is irrelevant to the analysis when we have well-defined probabilities to work with, ought we to expect it to affect a reasonable analysis of the data when the prior probabilities happen to be vague? No example strongly pointing toward an affirmative answer has yet been adduced.

Dr D. R. Cox: In the problem of the significance test, it seems to me that the Bayesian argument attains independence of the sampling rule by answering a somewhat different question from that we usually think about. Suppose that the null hypothesis is $\theta = 0$ and that we do some sort of optional stopping and end up with a very small $\bar{x}$ and a very large n. Now, as I understand the Bayesian point of view, the prior distribution must be fixed and independent of n; we have some prior probability at $\theta = 0$ and the remainder distributed in some way over the non-zero values of θ. If we ask the question 'Is θ zero or not?', we have in this case only two effective possibilities: either $\theta = 0$, or θ lies in a narrow band of width roughly $1/\sqrt{n}$ near 0. But the Bayes approach seems to me to have partly prejudged the issue by assigning very small prior probability to this latter band; it says that if θ is not zero, it is very unlikely to be in any particular narrow range. I think that putting in a prior distribution is causing us to answer a different question from 'Are or are not the data consistent with $\theta = 0$?' Now, of course, a further point often comes in that one says very small values of θ are practically unimportant and can be identified with the value of zero. That is a different issue. I think that the consideration of tail areas does enable us to deal with the question of consistency with a null hypothesis, without prejudging the issue by putting down a prior distribution that effectively excludes the possibility of a very small non-zero value of θ.

GOOD: A possible weakness in the use of the Bayes approach is in having a function which is smooth all the way to zero. It may be that the density function should tend to infinity, which in principle certainly comes closer to Dr Cox's case, so that after say, 10^{100} observations you would be able to say θ is very slightly different from zero. An assumption about a prior distribution that seems reasonable for a moderate size of experiment may not be advisable if you are going to do a very big experiment. You so to speak oversimplify because you know in advance that you are not going to do more than say a million experimental trials.

SAVAGE: What is essential to the Bayesian point of view, or approach, for the class of problem under discussion, is this. We believe that some prior distributions for the parameter θ will lead exclusively to beliefs and behaviour that you would regard as reasonable for the given situation. Since a tail area analysis, being in conflict with the likelihood principle, is not compatible with any prior distribution, and since the analogue of such an analysis is clearly contraindicated in exaggeratedly simple problems like simple dichotomy, we think it must be wrong.

To be sure, the particular kinds of prior distribution thus far mentioned in connection with hypothesis testing during this conference are not appropriate to all, or even to many, practical situations. Often, as you show, my actual prejudice against the parameter's lying near but not at the origin is less than a certain naïve model of my prior distribution would suggest, so that this model does not give a faithful image of my opinion in such a situation. To conclude from the inappropriateness of one kind of prior distribution that we should take seriously a procedure incompatible with all prior distributions seems to me to go further than is justified. The Bayesian theory does not yet have models of all, or even most, of the situations traditionally treated by hypothesis testing, but better analyses have not, to my knowledge, been demonstrated outside of Bayesian statistics.

BARNARD: I have been made to think further about this issue of the stopping rule since I first suggested that the stopping rule was irrelevant (Barnard, 1947a, b). This conclusion does not follow only from the subjective theory of probability; it seems to me that the stopping rule is irrelevant *in certain circumstances*. Since 1947 I have had the great benefit of a long correspondence – not many letters because they were not very frequent, but it went on over a long time – with Professor Bartlett, as a result of which I am considerably clearer than I was before. My feeling is that, as I indicated [on p. 42], we meet with two sorts of situation in applying statistics to data. One is where we want to have a single hypothesis with which to confront the data. Do they agree with this hypothesis or do they not? Now in that situation you cannot apply Bayes's theorem because you have not got any alternatives to think about and specify – not yet. I do not say they are not specifiable – they are not specified yet. And in that situation it seems to me the stopping rule is relevant.

In particular, suppose somebody sets out to demonstrate the existence of extrasensory perception and says 'I am going to go on until I get a one in ten thousand significance level'. Knowing that this is what he is setting out to do would lead you to adopt a different test criterion. What you would look at would not be the ratio of successes obtained, but how long it took him to obtain it. And you would have a very simple test of significance which said if it took you so long to achieve this increase in the score above the chance fraction, this is not at all strong evidence for E.S.P., it is very weak evidence. And the reversing of the choice of test criteria would I think overcome the difficulty.

This is the answer to the point Professor Savage makes; he says why use one method when you have vague knowledge, when you would use a quite different method when you have precise knowledge. It seems to me the answer is that you would use one method when you have precisely determined alternatives, with which you want to compare a given hypothesis, and you use another method when you do not have these alternatives.

SAVAGE: May I digress to say publicly that I learned the stopping-rule principle from Professor Barnard, in conversation in the summer of 1952. Frankly, I then thought it a scandal that anyone in the profession could advance an idea so patently wrong, even as today I can scarcely believe that some people resist an idea so patently right. I am particularly surprised to hear Professor Barnard say today that the stopping rule is irrelevant in certain circumstances only, for the argument he first gave in favour of the principle seems quite unaffected by the distinctions just discussed. The argument then was this: The design of a sequential experiment is, in the last analysis, what the experimenter actually intended to do. His intention is locked up inside his head and cannot be known to those who have to judge the experiment. Never having been comfortable with that argument, I am not advancing it myself. But if Professor Barnard still accepts it, how can he conclude that the stopping-rule principle is only sometimes valid?

BARNARD: If I may reply briefly to Professor Savage's question as to whether I still accept the argument I put to Professor Savage in 1952 (Barnard, 1947a), I would say that I do so in relation to the question then discussed, where it is a matter of choosing from among a number of simple statistical hypotheses. When it is a question of deciding whether an observed result is reasonably consistent or not with a single hypothesis, no simple statistical alternatives being specified, then the argument cannot be applied. I would not claim it as foresight so much as good fortune that on page 664 of the reference given I did imply that the likelihood-ratio argument would apply 'to all questions where the choice lies between a finite number of exclusive alternatives'; it is implicit that the alternatives here must be statistically specified.

SAVAGE: The question of imprecisely determined alternatives is provocative, but in the example of scores on a test for extrasensory perception, it seems to me that the alternatives are quite well specified. If the subject's mean score is not that of the null hypothesis, it is somewhat different, presumably higher. Something like section (*d*) of Dr Smith's contribution should apply, except that account should be taken of the fact that if there is any E.S.P. at all we expect it to be very small from general experience.

A valuable thing brought out by Professor Barnard's comments here and elsewhere is that often we are not only vague as to how our opinion is distributed over the possibilities but even vague as to what the possibilities are.

GOOD: What I call the device of imaginary results is relevant to the previous discussion. Usually we think of an argument from initial or prior probabilities and likelihoods to final probabilities and statistical inference. But if one takes the notion of consistency seriously it is just as legitimate to argue the other way. The words prior and posterior, or initial and final, might mislead one into forgetting this fact.

That is to say you can imagine certain possible final results of an experiment and then use Bayes's theorem in reverse in order to find out what your initial or prior judgements must be for the sake of consistency. For instance, to take a very simple example first, imagine an experiment in E.S.P. in which someone guesses forty consecutive cards correctly, each card having say five possible equally likely forms. You know that the likelihood of that on the null hypothesis, which is that there is no E.S.P. present and that the experiment is carried out honestly and accurately, etc., with no conscious or unconscious cheating, is 5^{-40}. Now if that happened would you or would you not believe that the man had power of extrasensory perception? If you would, then this tells you that the initial probability must exceed 5^{-40} and you have discovered something about your actual state of mind without actually doing the experiment. You merely imagine that this experiment could be performed. Likewise, in this discussion of the stopping rule, suppose you are estimating a probability near a half.