Falk, Raphael, 'The gene in search of an identity.' *Human Genetics*, *68* (1984) 195–204.
Falk, R. *Perception of Randomness.* Doctoral dissertation (in Hebrew with English abstract). Jerusalem: The Hebrew University, 1975.
Falk, R. 'The perception of randomness.' In *Proceeding of the Fifth International Conference for the Psychology of Mathematics Education* (pp. 222–229). Grenoble, France, 1981.
Fienberg, S. E. 'Randomization and social affairs: The 1970 draft lottery.' *Science*, *171* (1971) 255–261.
Gilovich, T., Vallone, R., and Tversky, A., 'The hot hand in basketball: On the misperception of random sequences.' *Cognitive Psychology*, *17* (1985) 295–314.
James, G. and James, R. C. (Eds.). *Mathematics Dictionary* (multilingual edition). Princeton: Van Nostrand, 1959.
Wagenaar, W. A. 'Generation of random sequences by human subjects: A critical survey of literature.' *Psychological Bulletin*, *77* (1972) 65–72.
Wagenaar, W. A. and Keren, G. B. 'Chance and luck are not the same.' *Journal of Behavioral Decision Making*, *1* (1988) 65–75.

Randomness is well enough understood to be misunderstood.[1]

Alexander Pollatsek
Department of Psychology.
Clifford Konold
Scientific Reasoning Research Institute, University of Massachusetts, Amherst.

In their article Ayton, Hunt and Wright address a number of issues that impinge on the concept of randomness. They appear to question not only the methodological soundness and general implications of research on 'misconceptions' in statistics, but also the soundness of aspects of statistical inference. We concentrate here on a few key issues about which we are in disagreement (we think) with the authors.

Rationality vs. Statistical Misconceptions

One of the basic purposes of their article is to rescue the lay person from charges of irrationality that have been based on subjects' non-normative performance on various statistical tasks. Their first line of defence against these charges involves a reiteration of Hume's injunction not to mistake empirical induction for a principle of logic. On the basis of the 'problem of induction,' Ayton *et al.* argue that because all inferences from real-world data rest on the same shaky ground, it is unfair to label only some of these (e.g., the belief in negative recency in coin flipping) as irrational. If the authors had limited their discussion to this strict sense of rationality, their argument would have been incontrovertible (though trivial). But they go on to discuss the rationality of naive conceptions of randomness based on the *adaptability* of those beliefs, and here their argument falters. We illustrate our point with an example.

Person *B* is assessing Person *C*'s beliefs about random events. *B* supplies a 'fair' coin, and they mutually agree on a shaking procedure that 'mixes' the coin well between flips. They flip the coin 100 times and at no point does *C* question the randomness of the results. They continue flipping until five straight heads occur, at which time *B* suggests that they wager on the outcome of the next flip. *C* quickly bets that a tail will occur, and after some bargaining agrees to give *B* 3 to 2 odds. Asked to justify this bet, *C* argues that a tail is much more likely than a head on the next flip because 'according to probability theory the coin needs to show a tail once in a while for it to be a fair coin.' *C*'s behavior may not be irrational in a strict sense. But the justification, even though understandable, is certainly incorrect and is not derived from probability theory. Furthermore, given that this specific bet is undertaken with the expectation of monetary gain, *C*'s behavior is not optimal, and therefore irrational in the adaptive sense of that term. If *C* persists in giving these odds in similar situations, *C* will lose money, Hume notwithstanding.

A discussion of whether actual experiments have indeed been convincing demonstrations of such irrationality is impossible here given space constraints. Any single experiment can always be given an alternative explanation. However, there are now numerous demonstrations that together are most parsimoniously explained by the hypothesis that a considerable percentage of people believe in, and even structure bets on, the 'gambler's fallacy.'

In addition to methodological problems in the research on conceptions of randomness, Ayton *et al.* claim that it is not irrational to believe in the gambler's fallacy because it is possible that in the real world, events that are loosely termed 'random' have more alternations built into them than perseverations. The general form of this argument is:

(1) Experience leads to the belief that x is true of situation A.
(2) Situation B looks like situation A.
(3) Assume x is true of B.

While inferences of this type are not unreasonable, they are frequently invalid. Indeed, one of the central points of the research of Tversky and Kahneman is that people develop heuristics that are reasonable in many contexts but apply them to situations for which they are inappropriate. Whether or not these particular overgeneralizations lead to dire consequences in the real world is indeed debatable. However, there should be no doubt that if people could be helped to distinguish between random and non-random events, and to reason normatively about the former, they could perform more optimally than they currently do. Given a populace that finally understood the impossibility of developing a successful betting system for randomly-generated outcomes, casinos and government-sponsored lotteries would undoubtedly feel the pinch. The market would also dry up for clever entrepreneurs who are selling randomizing devices to lottery players with the claim that numbers selected with this device have an improved chance of winning because they are generated in the same way as the winning numbers.

Inconsistencies in Tests of Randomness

In their attempt to reveal inconsistency and paradox in the testing of sequences for deviations from randomness, the authors present a confused montage of Bayesian and hypothesis-testing logic. As a result, it is often difficult to determine what they are saying. However, certain of their claims appear to be incorrect, or at least misleading.

For example, the authors appear to assert that Bayesian logic is helpless in allowing one to discriminate 'random' from 'non-random' hypotheses for a particular sequence without making a *specific* hypothesis about the non-random alternative. Their point may merely be that if one has *no* idea about the alternative hypothesis (e.g., the sequence was a result of God's whim), then Bayesian logic will get one nowhere. Their 'derivation' simply demonstrates that if one assumes the likelihood ratio to be one (by some strange equal-ignorance argument) then it follows (tautologically) that the datum will have no impact on the hypothesis. The relevance of this example to Bayesian logic is minimal.

They appear to be claiming something more, however, since they later assert that 'If we know a die as biased, but do not know *how*, we cannot revise the probabilities we would derive for a fair die' (p. 233). We think the authors are claiming here that Bayes' theorem is helpless not only in the case of 'no alternative hypothesis' but also in the case of fairly general hypotheses. To simplify the discussion, let us shift from a die to the case in which one is interested in testing whether a particular coin is 'fair' — i.e., if $p(\mathrm{H}) = 0.5$, assuming independence. In a Bayesian framework, there are standard calculations for posterior odds (given a sequence of observations) that avoid the necessity of positing a particular alternative hypothesis. For example, one merely has to posit some *a priori* distribution (e.g., rectangular on the interval [0,1]) on the parameter p of obtaining a head to arrive at a posterior distribution for the probability of heads. While the posterior distribution on p will depend on the particular prior distribution selected, a strength of the Bayesian method is that this distribution will converge on the observed frequency given large amounts of data, assuming almost any *a priori* distribution.

Their treatment of hypothesis testing is even more confusing. They repeatedly point out, for example, that regardless of conclusions drawn about the randomness of a particular sequence, 'it is not possible to verify beyond all doubt that any given real sequence is, or is not, random.' (p. 227). They seem to regard this truism about Type I and Type II errors as a telling argument against the validity of tests of randomness *per se* rather than as a good example of what Hume was talking about. They also find it paradoxical that we cannot put together a random series of length n by stringing together m sequences of length n/m, each of which has passed some randomness test (p. 230). Of course, the reason that we cannot is because in the course of testing sequences we would eliminate (through Type I error) a certain percentage of the random sequences of length n/m. More to the point, a random sequence is not one in which there is a 'complete absence of every possible type of pattern' (p. 232), but one which was *generated* via a process of random selection. Does this mean that the concept of randomness is ill-formed because we have no generating mechanism that is truly random? Not unless we also regard the concept of a circle problematic because we have no method of generating a perfect one — that however precisely a particular circle may be constructed, it can always be shown under closer scrutiny to deviate from circularity.

Nor is it as problematic as the authors suggest (p. 230) to reject the hypothesis that a particular coin (or a person calling out Hs and Ts in an attempt to mimic a coin) is 'fair'. *A priori*, the two most common deviations one would be on guard against are (a) that p is not 0.5 and (b) first-order sequential dependency. Tests can be conducted with modest samples in which large deviations from $p = 0.5$ and independence can usually be detected. Of course, as the authors point out, peculiar higher-order sequential dependencies are difficult to detect. The practical implications of this difficulty or how it relates to the research on the gambler's fallacy escape us.

The authors do suggest in conclusion that there might be some circumstances in which there are practical consequences for seeing patterns where there are none. In this connection they mention the belief among basketball enthusiasts in the 'hot hand.' We would quickly add that guarding against these 'practical consequences' is the major purpose for using inferential statistics in the social sciences, where a large portion of the results could parsimoniously be explained as coming from a random device, and it is worthwhile to establish for a certain set of data that at least that hypothesis can reasonably be rejected. We think the tools generated for this purpose are quite useful and have succeeded in discriminating 'non-random' data from data that, for the purposes of interpretation, may as well be 'random.' And to find that those who are trained in the use of statistical methods frequently do no better than novices on various statistical-reasoning tasks ought to produce some concern, if not for the cause of rationality in everyday decision-making, then certainly for the cause of whatever it is we hope to foster through formal education.

NOTE

1. The preparation of this article was supported by grant MDR-8954626 from the National Science Foundation.

Randomness and Randomizers: Maybe the Problem is not so Big.

Willem A. Wagenaar
Leiden University, The Netherlands.

The central theme in the challenging paper by Ayton, Hunt, and Wright is that we will never find out whether people are good randomizers and good detectors of randomness, if we do not generate hypotheses about what else they might be. I agree with this tenet, but I don't think this poses a serious problem: there are some useful and plausible hypotheses available. It is even the case that testing some or even most of these alternative hypotheses render the discussion about what randomness really is, or how it is measured, somewhat superfluous. For sake of simplicity I will limit the discussion to the generation paradigm, but the arguments apply equally well to problems of induction and detection of randomness.

Properties, not sequences

When studying randomization behaviour, we should be interested, not in the generated sequences as such, but in properties of those sequences. Take as an example Mittenecker's (1953) hypothesis that subjects randomize by drawing without replacement. Thus, they would mimic throwing dice by placing successive permutations of the numbers 1 to 6 in a long sequence. Such an hypothesis can be tested by considering some of its consequences: no runs longer than 2, runs of 2 occur once in 36 elements, runs of 2 have a minimum distance of four other elements in between, the maximum number of other elements before repetition of an element (= maximum recurrence distance) is 10, the distribution of recurrence distances is triangular with a peak at 5, etc. Acceptance or rejection of the model depends on such properties of the generated sequences. For instance, a run of three identical elements cannot be produced by Mittenecker's process, and would, if it occurs, refute the model. That argument is not weakened by the fact that all specific sequences are equally likely to be produced by a random generator. I admit that it is not perfectly clear how much discrepancy between predicted and observed properties can be tolerated, but that problem is common to the testing of all behavioural theories.

Mittenecker's hypothesis is not very enlightening, because it involves permutations produced by a randomizer, and it is not explained how this randomizer works. However, rejecting Mittenecker's hypothesis for its 'homunculus' character does not depend on defining a 'standard' of randomness, or on application of a randomness test.

The acceptance or rejection of a model can be (but need not be) guided by the Bayesian principle of accepting the model that is most likely in view of the data. This would necessitate a specification of some properties of sequences produced by a process of random selection. For instance, the principle that each element is generated independently of previous elements implies that the recurrence distribution is described by a decreasing, negatively accelerated, geometric function. Testing the data against the two hypotheses of a triangular or a geometric recurrence distribution does not conflict with the principle that any response sequence is equally likely to be produced by a random generator. The equal likelihood holds for response sequences, not for shapes of recurrence distributions.

Properties of generators, not of sequences

Mittenecker's hypothesis differs from a random generator with respect to replacement of the drawn elements, and this affects the independence property. Therefore any statistic, other than the recurrence distribution, that reflects the independence property, provides an opportunity for comparing the two competing hypotheses.

The definition of randomness is not an obstacle here. The real problem is that randomness is in reality a property of a generator, not of its products. It must be admitted that inferring properties of generators on the basis of their products will always be problematic. Hypothesis testing in the Neyman-Pearson tradition is logically flawed, and Bayesian hypothesis testing depends on the ingenuity of the scientist imagining alternative hypotheses. It would be useful to establish properties of generators without the mediation of their products. An example of such a property is the independent selection