PSYCHOLOGICAL BULLETIN
Vol. 54, No. 4, 1957

THREE CRITERIA FOR THE USE OF ONE-TAILED TESTS

HERBERT D. KIMMEL
University of Southern California

Examination of the recent literature on the question of when to use one-tailed tests of significance in psychological research reveals a state of unresolved disagreement. A variety of differing opinions (**1**, **2**, **5**, **7**, **8**, **9**, **10**, pp. 62–63) have been presented, ranging from Burke's (**2**) exhortation that psychologists should never report one-tailed tests in the public literature to Jones' (**8**) statement that we may not only do so, but, in certain instances, we will be in error if we fail to do so.

It is by no means necessary for psychologists to agree on all matters of importance to them. Disagreement regarding methodological considerations, however, especially when they bear on how and when propositions shall be accepted as true or rejected as false, should not be permitted to persist indefinitely. The argument is not settled by noting, as Burke (**2**) does, that the increased use of one-tailed tests may result in the one-tailers scoring a sociological victory almost before the controversy has begun. Actually, this observation by Burke does not coincide completely with the fact that many responsible investigators have continued to employ two-tailed tests (in situations calling for one-tailed tests according to Jones' view) long after the opening of the one-tailed avenue.[1]

In attempting to arrive at a set of acceptable criteria for the use of one-tailed tests, it is important to note that the argument is not one of mathematical statistics but primarily one of experimental logic. Burke and Jones would agree that one-tailed tests should be used to test one-tailed hypotheses; their disagreement concerns when one-tailed hypotheses should and should not be made.

Before proceeding to the proposed criteria, it would be of value to consider the difference between one- and two-tailed hypotheses from a viewpoint that has not been stressed by previous writers. All concerned agree that a given mean difference in the hypothesized direction is "more significant"[2] under a one-tailed hypothesis (in the correct direction) than under a two-tailed hypothesis. This is due to the fact that there are exactly twice as many chances of

[1] An example of an experiment with an explicit directional hypothesis, but employing a two-tailed test, is reported by Davitz (**3**). This experimenter reasoned that the injection of tetraethylammonium prior to extinction trials would inhibit the punishing effect of the emotional response under study and, consequently, would result in faster extinction in the experimental animals than in a placebo-injected control group. Instead, Davitz found that the experimental group extinguished slower than the control group, the difference in mean number of trials being significant at the 5 per cent level using a two-tailed test. A one-tailed hypothesis in this experiment (as would have been urged by Jones) would have made it impossible to evaluate the significance of the obtained difference. A study by Hilgard et al. (**6**), on the other hand, stated a one-tailed hypothesis in a situation in which a difference in the unpredicted direction could have been predicted with as much justification on the basis of previous work. They obtained a difference in their predicted direction that was significant at the 5 per cent level using a one-tailed test. Their rejection of the null hypothesis on the basis of the difference they obtained is the equivalent of loosening the conventional 5 and 1 per cent standards.

[2] That is to say, by chance, the unidirectional event is half as probable as the bidirectional; thus its occurrence, being half as likely, is twice as significant.

committing a type 1 error, with a given mean difference, under a two-tailed hypothesis. The important consideration is that this gain does not accrue without concomitant loss. Even psychology has its law of conservation of energy.

The price that is paid in return for the increased power of one-tailed tests over two-tailed tests stems from the fact that two-tailed null hypotheses are actually more specific than their one-tailed counterparts. A two-tailed null hypothesis can be rejected by a large observed difference in either direction but a one-tailed null hypothesis cannot be rejected by a difference in the unpredicted direction, no matter how large this difference may be. This means that an experimenter using a one-tailed hypothesis *cannot* conclude that an extreme difference in the unpredicted direction is reliably different from zero difference. This limitation cannot be shrugged off by the comment, "We have no interest in a difference in the opposite direction." Scientists are interested in empirical fact regardless of its relationship to their preconceptions.

The meaning of this limitation is exemplified even in applied studies; e.g., those intended to answer the question whether a new product is "better" than the current product. It would be desirable to be able to conclude that the new product is not only "not better" (which is all that failure to reject a one-tailed null hypothesis permits[3]), but, in fact, "poorer." The decision not to market the proposed new product would follow from either conclusion, it is true, but the additional information available as a result of rejecting a two-tailed null hypothesis from the unexpected side could very well indicate a course of behavior quite different from that indicated by the mere inability to reject a specific one-tailed null hypothesis.

It is hoped that the following criteria will be acceptable to psychological investigators as a group and will be adopted conventionally as a guide. The ultimate consequence of our present state of ambiguity on this matter can only be confusion and subsequent retrogression to a more primitive level of scientific communication and understanding.

Criteria for the Use of One-Tailed Tests

1. Use a one-tailed test when a difference in the unpredicted direction, while possible, would be psychologically meaningless. An example of this situation might be found in the comparison of experimental and control groups on a skilled task for which only the experimental group has received appropriate training. The experiment would have to be designed in such a way as to eliminate all known conditions that could produce opposite results (e.g., not testing immediately after training to avoid fatigue effects, not testing too long after training to avoid memory loss effects, etc.). Since a difference in the unpredicted direction will have been declared beforehand to have no possible meaning (in terms of previous data and present operations) one-tailed hypotheses could not undergo metamorphosis into two-tailed hypotheses to permit testing the significance of differences in the unpredicted direction.

2. Use a one-tailed test when re-

[3] As Fisher (4) has pointed out, an experimenter never "accepts" the null hypothesis, he merely *fails to reject it on the basis of his* data. This is one reason why the null hypothesis in a particular experiment should be stated as specifically as possible.

sults in the unpredicted direction will, under no conditions, be used to determine a course of behavior different in any way from that determined by no difference at all. This situation is exemplified by the applied study discussed above, in which a new product is compared with one already on the market.

3. Use a one-tailed test when a directional hypothesis is deducible from psychological theory but results in the opposite direction are not deducible from coexisting psychological theory. If results in the opposite direction are explainable in terms of the constructs of existing theory, no matter how divergent from the experimenter's theoretical orientation this theory may be, the statistical hypothesis must be stated in a way that permits evaluation of opposite results. If this criterion were not already implicitly accepted by psychologists, crucial experiments could never be performed.

It should be apparent that the three criteria stated above are actually slightly differing reflections of the same underlying precept. Neither the ethical nor the logical decisions of individual scientists can be prescribed beforehand by any set of standards, no matter how all-pervasive these standards may seem at a given moment. The three criteria proposed above, however, are offered as temporary guideposts until such time as a new set of temporary criteria supersede them. Proponents of one-tailed tests, such as Jones (**7**, **8**), cannot complain that the use of these criteria will reduce the number of one-tailed tests to near zero, without admitting that these tests have been misused in the past. Opponents of one-tailed tests, such as Burke (**1**, **2**), should welcome this attempt to limit the use of one-tailed tests to those infrequent situations provided for by the proposed criteria.

REFERENCES

1. BURKE, C. J. A brief note on one-tailed tests. *Psychol. Bull.*, 1953, **50**, 384–387.
2. BURKE, C. J. Further remarks on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 587–590.
3. DAVITZ, J. R. Decreased autonomic functioning and extinction of a conditioned emotional response. *J. comp. physiol. Psychol.*, 1953, **46**, 311–313.
4. FISHER, R. A. *The design of experiments.* London: Oliver and Boyd, 1947.
5. HICK, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, **59**, 316–318.
6. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
7. JONES, L. V. Tests of hypotheses: One-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, **49**, 43–46.
8. JONES, L. V. A rejoinder on one-tailed tests. *Psychol. Bull.*, 1954, **51**, 585–586.
9. MARKS, M. R. Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Rev.*, 1951, **58**, 179–184.
10. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

Received February 21, 1957.