

TESTING TWO-STATE THEORIES WITH OPERATING CHARACTERISTICS AND A POSTERIORI PROBABILITIES¹

WAYNE A. WICKELGREN

Massachusetts Institute of Technology

The use of operating characteristics (OCs) and a posteriori probability functions to test 2-state theories of perception or memory is discussed. Such tests require the validity of certain additional assumptions about the operation of the decision system which maps the states of the perceptual or memory system into responses. In the case of OCs generated by the payoff method using binary responses, the required decision-making assumption is one specified by Luce. His 2-state decision rule has been tested by means of a decision-making experiment and shown to be valid. Thus, the binary-response OC generated by the payoff method can be used to both accept and reject the 2-state hypothesis in any situation. Questions are raised with regard to other methods of generating OCs concerning their suitability for testing the 2-state hypothesis. Confidence judgments (rating methods) are found to be less useful for testing 2-state theories, whether the results are analyzed by OCs or a posteriori probability functions.

In recent years there has been considerable interest in two-state (all-or-none, threshold) theories of both perception and memory in a variety of situations. The alternatives to a two-state theory are many and varied. On the one hand, the perceptual or memory system might have a small finite number of states greater than two. These states might be ordered, partially ordered, unordered, have a metric defined on them, etc. On the other hand, the number of states might be so large and well behaved that the continuous real number system might be a good approximation.

This paper is concerned with the testing of two-state theories in what might be called two-alternative absolute judgment tasks, namely, tasks in which one of two alternative test stimuli (Y or N) is presented, to which the subject must make one of two alternative responses (y or n). For the purpose of this paper it makes no difference whether the task is signal detection (present vs. absent), absolute pitch judgment (high vs. low), absolute length judgment (long vs. short),

¹ This work was supported primarily by Grant MH 08890-03 from the National Institute of Mental Health, United States Public Health Service. Further aid was received from National Aeronautics and Space Administration Grant NsG 496, to Hans-Lukas Teuber.

recognition memory (same vs. different from a previous stimulus), or whatever. Furthermore, for the present discussion it makes little difference what alternative is envisaged to a two-state theory. The real line of signal detection theory is the most frequent alternative.

The following is the basic theoretical framework within which the two-state versus multistate (or continuous) question is asked. The organism is assumed to be analyzed into two subsystems, a sensory or memory (SM) system and a decision (D) system. Either of the two alternative stimuli enters the SM system, which produces an output to the D system, which in turn makes a y or n response based on this information. This paper is concerned with testing the hypothesis that the output of the SM system to the D system has one of only two values, 0 or 1. To provide a clear intuitive picture of the situation, let us suppose that Stimulus Y , State 1, and Response y tend to be associated with each other, and Stimulus N , State 0, and Response n tend to be associated with each other. The most general two-state model of such a system is

$$P = Pr(y|Y) = pf + (1 - p)g$$
$$Q = Pr(y|N) = qf + (1 - q)g$$

where p and q are the probabilities that the

SM system is providing 1 output to the D system in situations where Y or N stimuli are being presented, respectively, and where f and g are the probabilities that the D system responds y when it receives 1 or 0 input, respectively.

Presumably, p and q are functions of the two stimulus conditions, which we shall assume to remain constant throughout our experiment. On the other hand, f and g are response biases, which we shall assume to be varied during the experiment by some procedure such as manipulating the 2×2 payoff matrix for y and n responses given Y and N stimuli. It is absolutely essential that the manipulation of payoffs not affect the SM system (i.e., p and q). However, if there is any reason to doubt this assumption in the usual way of running two-alternative absolute judgment experiments, its validity can be virtually guaranteed by presenting the payoff matrix only *after* the stimulus and varying the matrix from trial to trial. The disadvantages that one might imagine for this method are shown later to be absent for the purposes of testing a two-state theory.

No testable predictions have yet been derived from this completely general two-state theory. In particular, as pointed out by Broadbent (1966), it does not predict that the operating characteristic (OC) obtained by plotting P against Q for various payoff matrices will consist of two straight lines intersecting at the point q, p . The prediction that the OC will consist of two straight lines derives from a *special case* of the general two-state theory in which a strong assumption is made about the admissible pairs of values of the response biases f and g , namely, that $f < 1$ implies that $g = 0$, and $g > 0$ implies that $f = 1$. In other words, if we start with a condition in which the subject is overwhelmingly biased to respond "n" ($f = 0$ and $g = 0$), and begin to increase his bias to respond "y," the subject is assumed to first increase the value of f from 0 to 1 while leaving g at 0. When $f = 1$ and $g = 0$, further increases in the bias to say "y" are due to increases in g from 0 to 1, while f remains constant at 1. This is Luce's (1963) two-state threshold theory, and it is this theory

that predicts OCs consisting of two straight lines, the lower limb going from 0,0 to q, p and the upper limb going from q, p to 1,1.

The equation for the lower limb is obtained by setting $g = 0$, solving for f , and equating to yield $P = (p/q)Q$. The equation for the upper limb is obtained by setting $f = 1$, solving for g , and equating to yield

$$P = [(1-p)/(1-q)]Q + [(p-q)/(1-q)]$$

In the absence of any knowledge of whether the decision-making assumptions of Luce's (1963) theory are correct, it is impossible to decide definitely whether the SM system has only two possible states (more correctly, provides only one of two outputs to the D system) for any particular task. Thus, all previous experiments that have yielded OCs consistent with a continuous or multistate theory, and inconsistent with Luce's two-state theory, are *either* inconsistent with the two-state SM assumption *or* with Luce's two-state decision-making assumption. Without independent data on the decision-making assumption, we do not know which assumption is at fault. Of course, in those less frequent cases where the OC closely approximates two straight lines, we can be rather sure of both the two-state SM assumption and Luce's two-state decision-making assumption. If one had independent evidence for the validity of Luce's two-state decision-making assumption, then one could be equally sure that the two-state SM assumption was invalid wherever OCs were not well fit by two straight lines. Since the latter is probably the more usual case, such evidence is quite important for the definitive rejection of the two-state assumption in a variety of perceptual and memory tasks.

Therefore, a decision-making experiment was done to test Luce's rule under two instructional sets, one designed to mimic a perceptual (detection) task as closely as possible, and the other a recognition-memory task. An artificial (experimentally controlled) two-state perceiver or memorizer was used in these decision-making experiments. The artificial perceiver was only partially accurate, and subjects knew the accuracy only within a rather broad range (60-90%). Response

bias was manipulated by payoff matrices. The results completely confirmed the validity of Luce's two-state decision rule (Wickelgren, 1967).

Although every effort was made to provide the dichotomous information to the subject's decision system in a context similar to that of a detection or memory experiment, obviously the subject could tell the difference. Thus, one could always argue that the validity of Luce's rule in the decision-making experiment was not absolute proof of its validity in a different situation, regardless of the degree of similarity. Nevertheless, the experiment, in conjunction with one's intuitive confidence in the validity of Luce's rule, does suggest that we need have little concern about the validity of Luce's rule in testing the two-state SM assumption by means of OCs when the OCs are generated by varying payoff matrices. However, in future tests of the two-state SM assumption using OCs, it might be wise to run the subjects in a short decision-making experiment of the present type to be sure that these subjects follow Luce's rule.

Assuming that Luce's rule is valid for all subjects in all situations where response bias is manipulated by payoffs, it would be nice to be able to review the literature in some area and produce a variety of instances where the evidence allowed one to definitely accept or reject the two-state SM assumption. Surprisingly enough, there are few published studies in detection, or other areas, where OCs were generated by payoff matrices (or any other binary-decision method, for that matter). As Luce (1963) has pointed out, most existing binary-decision OCs for signal detection are composed of points with variability greater than the binomial, and there is one OC for a subject in the Swets, Tanner, and Birdsall (1961) study of visual detection that is obviously better fit by a two-state threshold theory than by signal detection theory.

However, all other subjects cited by Luce (1963) show a systematic deviation from the two-straight-line OC predicted by Luce's two-state threshold theory, namely, rounded corners. This deviation is exactly what would be

predicted by signal detection theory. Furthermore, the same sort of deviations were found in the recent study by Galanter and Holman (1967). Luce's (1963) explanation of the systematic deviation really predicts the absence of points at x and y coordinates of the corner (a hole at the corner), not the systematic occurrence of points below the sharp corner (rounded corner). So far as one can tell from published data, it is a rounded corner, not a hole at the corner, which is almost always found.

One can explain the rounded corners in a manner consistent with Luce's two-state theory whenever it seems reasonable to assume that subjects have not reached stable response biases and are shifting from one limb of the OC to the other, with the payoff matrix held constant. However, if subjects are given extensive practice with feedback and then feedback is discontinued for the experimental session, the decision-making experiment reported previously indicates that there would be no shifting from one limb to the other.

Nevertheless, at the present time there are not enough OCs obtained by varying payoff matrices (Swets et al., 1961, and Galanter and Holman, 1967, are the only studies to my knowledge) to draw a definite conclusion about the validity of the two-state SM assumption for any task. Naturally, rejection of the two-state assumption for one type of task would not reject the assumption for a task where a different sensory or memory system is being employed.

VARYING A PRIORI PROBABILITIES OR RESPONSE-FREQUENCY INSTRUCTIONS

All of the previous discussion on testing the two-state hypothesis assumed that OCs were obtained by varying the payoff matrix. Whether Luce's two-state decision rule holds when the bias to respond y is manipulated by varying the a priori probability of presenting Y or by instructing the subject to respond y or n incorrectly no more than $x\%$ of the time (Neyman-Pearson criterion) is impossible to say on the basis of present evidence. Decision-making experiments that seem a priori to be psychologically equivalent

to the decision-making aspect of an absolute-judgment task are more difficult to construct in these cases, but if someone felt there was a need for another method of generating OCs to test the two-state hypothesis, a suitable decision-making experiment could probably be devised.

All three of these binary-choice methods are supposed to yield equivalent OCs, though the evidence for this is sparse (for exceptions, see Galanter & Holman, 1967; Green & Swets, 1966, pp. 88-89). Assuming this is true, there would be little incentive to test the two-state assumption with more than one method of generating the OC.

CONFIDENCE RATINGS

If a subject truly had only two states in his SM system and were faced with the problem of choosing one of six rating responses, the subject would either think the experiment or the experimenter was pretty stupid, or else that there was something wrong with him or his understanding of the task. In either case, it is not clear what his decision rule would be. A decision-making experiment to determine how he assigns confidence ratings with a two-state SM system might show that the confidence rating procedure disrupts the subject's adherence to Luce's two-state decision rule for the y - n decision. Even if the y - n decision conformed to Luce's rule, there are many possible rules the subject might use for choosing a confidence rating, and no one rule stands out as being appreciably more likely than the others.

Certainly, as Larkin (1965) has pointed out, if the SM system had only two states, it is very unlikely that the rating OC would be identical to the binary-choice OC.

It is somewhat more likely that the rating OC would consist of two straight lines, but this prediction is far from being as plausible as the comparable prediction for binary-choice OCs. For example, the most plausible way in which a rating OC for a two-state SM system might turn out to be two straight lines is the way assumed by Nachmias and Steinman (1963) and by Broadbent (1966). The subject first makes a binary decision, presumably the y - n decision, but the decision could be between the two response classes on

either side of any cutting point on the ordered rating scale. Having made the binary decision as to which of the two "equivalence" classes of responses to choose, he selects the particular response from that set at random. This sort of decision procedure would produce an OC composed of two straight lines, although it should be noted that unless the binary decision were at the equal-bias point ($f = 1$, $g = 0$), this two-straight-line rating OC would not coincide with the two-straight-line binary-choice OC.

Furthermore, if the binary decision is not at the equal-bias point, then it seems very likely that the assignment of confidence ratings would be nonrandom on the elongated limb of the OC, causing that limb to be a convex-up curve rather than a straight line. This could easily make the entire OC appear curved in the same manner. The reason for the nonrandom assignment of confidence ratings is easy to see. Assume the subject is more biased to respond y than n . Thus, the lower limb is the elongated limb. In this case, assuming Luce's two-state decision rule, the subject is always responding y to 1 output from the SM system, but is sometimes responding y to 0 output. However, in such a case, it would seem quite reasonable for him to assign, on the average, greater confidence in the 1 output cases than in the 0 output cases. This would produce a convex-up curved limb of the OC.

If the subject is not following Luce's decision rule when he has to assign confidence ratings, but using the completely general two-state decision rule, then the entire OC can be curved.

For exactly the same reasons that the rating OC can be curved with an underlying two-state SM system, the a posteriori probability function need not be composed of two horizontal line segments. Thus, for example, the multivalued a posteriori probability functions found for visual detection by Nachmias and Steinman (1963) and for recognition memory (particularly of single-digit numbers) by Norman and Wickelgren (1965) do not rule out two-state SM systems for these tasks.

Similarly, in n -alternative multiple-choice

tasks ($n > 2$) such as perceptual identification or a recall test of memory, it does not disprove the two-state SM assumption to show that greater confidence in one's response is associated with a smooth monotonically increasing probability that that response is correct.

In the first place, it is not clear what the two-state assumption should be in these cases. For example, in a paired-associate recall experiment, one imagines that the two-state assumption could be that the associations from the stimulus item to *each* of the possible response items are either learned or unlearned. If, as seems reasonable, the D system receives information about the state of several or all of the associations to the stimulus item, then while each association has only two possible states, the SM system has more than two states and is, in fact, providing an n -dimensional vector of inputs to the D system.

Of course one can assume that incorrect associations are never in the learned state. In this case, the SM system has only two states, $(0, 0, \dots, 0)$ and $(1, 0, \dots, 0)$. However, a smoothly increasing probability of correct response with increasing confidence would not disprove even this very special case of the two-state memory theory. To see this, one must consider all the logical possibilities for the operation of the D system in making a choice among the responses and the confidences, given its two-valued input. Let us take the simplest case first. Assume that when the input is $1, 0, \dots, 0$, the subject always chooses the correct response, and when the input is $0, 0, \dots, 0$, the subject selects randomly from the n alternatives. What can we assume about the assignment of confidence ratings? We can assume that in the former case the subject chooses from a set of high confidences and in the latter case chooses from a nonoverlapping set of lower confidences. This combination of SM and D assumptions would produce a recall a posteriori probability function composed of two horizontal straight-line segments.

However, virtually any other assumption about how the D system assigns confidences will not lead to this simple prediction, and the reasonable alternative assumptions about

the D system would produce smoothly increasing recall a posteriori probability functions even with the simplest two-state SM system. For example, even with the simple choice-response decision rule mentioned previously, if the two sets of confidence levels overlap, the recall a posteriori function will have more than two values.

If subjects do not always choose the correct response when the input to the D system is $1, 0, \dots, 0$, then one is very likely not to get the two-valued recall a posteriori probability function.

Finally, if incorrect associations can be in the learned state, the chances of obtaining a two-valued recall a posteriori probability function are nil.

Thus, the smoothly increasing recall a posteriori probability functions of Suboski, Pappas, and Murray (1966a, 1966b) do not rule out a two-state theory of the memory system.

A completely equivalent argument can be given for perceptual identification of one of n alternative stimuli. In both recall and perceptual identification, definite decisions about two-state theories cannot be made using confidence judgments without knowledge of how the D system works to produce both the choice response and the confidence rating. Lacking such knowledge of the D system, we cannot rule out a two-state theory of the SM system on the basis of confidence judgments.

Returning to the two-alternative case again, Broadbent (1966) has argued that the equivalence of binary-response and rating OCs allows us to use one just as well as the other to test the two-state theory of the SM system. There is some merit in this point, but it does require us to assume that because the two OCs are equivalent in some situations, they are equivalent in all situations. Given the results of the decision-making experiment reported in this paper, it is undoubtedly the case that wherever convex-up OCs have been obtained by the binary-response payoff method the two-state assumption is invalid. In such cases, one would expect the rating OC to be equivalent to the binary-response OC.

However, in a new situation, the two-state assumption might be valid. It is precisely in these situations where one would *not* expect the rating OC to be equivalent to the binary-response OC. Thus, it seems dangerous to rely on the rating OC alone in testing the two-state hypothesis, although the rating method might be used in addition to the payoff method to provide some additional support for the interpretation which must be based primarily on the binary-response results.

To use the rating method to test the two-state SM assumption, we need to know whether subjects follow Luce's two-state decision rule for their binary choices when they also have to give confidence ratings. If they do, we need to know whether confidences are assigned randomly after the binary choice is made. There is at least one case where random assignment of the confidence levels is assured, provided subjects are following Luce's rule, namely, the case where subjects are at the equal-bias point for their binary choices. However, to assure equal bias, as well as adherence to Luce's rule, is not easy, and a method that requires both is less useful for testing the two-state assumption than a method that requires only adherence to Luce's rule.

All this should not be taken to mean that the payoff method is, in general, better than the rating method for generating operating characteristics. All that has been suggested is that the payoff method is better than the rating method for testing the two-state assumption. If one already has evidence against the two-state assumption in some particular situations, or is willing to assume a large number of states without direct proof, then

one can give arguments (besides efficiency) for preferring the rating method to the payoff method (Wickelgren, 1968).

REFERENCES

- BROADBENT, D. E. Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America*, 1966, 40, 244-245.
- GALANTER, E., & HOLMAN, G. L. Some invariances of the isosensitivity function and their implications for the utility function of money. *Journal of Experimental Psychology*, 1967, 73, 333-339.
- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- LARKIN, W. D. Rating scales in detection experiments. *Journal of the Acoustical Society of America*, 1965, 37, 748-749.
- LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, 70, 61-79.
- NACHMIAS, J., & STEINMAN, R. M. Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, 1963, 53, 1206-1213.
- NORMAN, D. A., & WICKELGREN, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 1965, 70, 479-489.
- SUBOSKI, M. D., PAPPAS, B. A., & MURRAY, D. J. Confidence ratings in recall paired-associates learning. *Psychonomic Science*, 1966, 5, 147-148. (a)
- SUBOSKI, M. D., PAPPAS, B. A., & MURRAY, D. J. Confidence ratings in recall paired-associates: The RTT paradigm. *Psychonomic Science*, 1966, 5, 315-316. (b)
- SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, 68, 301-340.
- WICKELGREN, W. A. A test of Luce's two-state decision rule. *Psychonomic Science*, 1967, 9, 91-92.
- WICKELGREN, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 1968, in press.

(Received March 30, 1967)