Psychological Review
1994, Vol. 101, No. 1, 166–171

0033-295X/94/$3.00

Comments on Batchelder and Riefer's Multinomial Model for Source Monitoring

R. A. Kinchla

Judging not only whether a test item had been presented earlier but also its source (e.g., presented acoustically or visually, in red or green, in List 1 or List 2, and so on) is a form of recognition memory called *source monitoring* (or sometimes *reality monitoring* or *list discrimination*). There is a clear disparity between Batchelder and Riefer's (1990) model of source monitoring and a large body of data on recognition memory. Thus, despite the elaborate statistical techniques they developed for estimating parameters of their model, these estimates may not represent psychologically valid measures of source monitoring. It is suggested that models that are based on multidimensional statistical decision theory may be more appropriate.

Batchelder and Riefer (1990) proposed a multinomial model for memory studies of *source monitoring*. These are studies of recognition memory in which subjects are not only asked to distinguish old from new items but also to distinguish the source of the old items, for example, in which of two lists an old item originally appeared. They stated that estimating parameters of their model is "a way to separately measure the overall detectability of old items from the ability to discriminate the source of the items" (Batchelder & Riefer, 1990, p. 550).

I assert here that Batchelder and Riefer (1990) did not adequately explain the relationship of their model to a previously discredited model of signal detection and recognition memory. Although they briefly acknowledged that their model was built on a particular high-threshold (HT) model of signal detection, they did so in a way that may have led many readers to assume it was only a convenient variant of the Green and Swets (1966) statistical decision theory (SDT) model of signal detection (Batchelder & Riefer, 1990):

> Models based on auditory signal-detection principles (Green & Swets, 1966) have facilitated the analysis of recognition memory paradigms, so our goal in this section is to extend signal-detection notions to model source monitoring. . . . [which] is analogous to a standard yes–no signal-detection task, except that the "signal" is one of two sources of items. . . . There are several explicit ways to create signal-detection models [for such tasks] . . . but in the interest of simplicity, we decided to build our source-monitoring models on high-threshold versions of the corresponding signal-detection models (e.g., see Laming, 1973, chap. 6). (p. 550–551)

In fact, rather than a convenient variant of the SDT model, the particular HT model Batchelder and Riefer (1990) referred to is the old Fechnerian view of a threshold that Green and Swets (1966) specifically rejected. Batchelder and Riefer did not directly acknowledge this, nor did they mention the variety of experimental evidence on which this rejection was based, they simply stated that "High-threshold models for signal detection are probably oversimplified if viewed as a precise theory of auditory signal detection; however, they are known to provide useful measurement tools in recognition memory" (p. 551).

Actually there is considerable evidence that the particular HT model Batchelder and Reifer (1990) used is a very poor model for both auditory detection and recognition memory.

Shortcomings of the Fechnerian HT Model

Before considering the multinomial model, it is useful to briefly characterize the Fechnerian HT model and the experimental evidence against its use as either a model of auditory detection or of recognition memory.

Fechnerian HT Model

Blackwell (1963) formalized the old Fechnerian view of a sensory threshold with the stochastic process represented in Figure 1. On signal (S_1) trials a subject either (a) detects the signal (with probability α) and responds yes (R_1) or (b) fails to detect the signal (with probability $1 - \alpha$), then either guesses yes (with probability β) or no (R_0). On no-signal (S_0) trials the subject never detects a signal and again either responds yes (with probability β) or no otherwise. (Hence, the name high-threshold model: A nonsignal never exceeds the detect threshold.)

Performance on a yes–no signal-detection task can be summarized by a subject's tendencies to respond yes on signal and on nonsignal trials: the so-called hit and false-alarm rates, respectively. For the HT model, these correspond to the following theoretical probabilities that are based on the trees in Figure 1:

$$P(R_1 \mid S_1) = \alpha + (1 - \alpha)\beta \tag{1}$$

and

$$P(R_1 \mid S_0) = \beta. \tag{2}$$

So, by substitution in Equation 1 using Equation 2,

$$P(R_1 \mid S_1) = \alpha + (1 - \alpha)P(R_1 \mid S_0). \tag{3}$$

Equation 3 defines a linear *operating characteristic* (OC) for each value of α as illustrated in Figure 2. Each linear function

Correspondence concerning this article should be addressed to R. A. Kinchla, Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08544-1010.

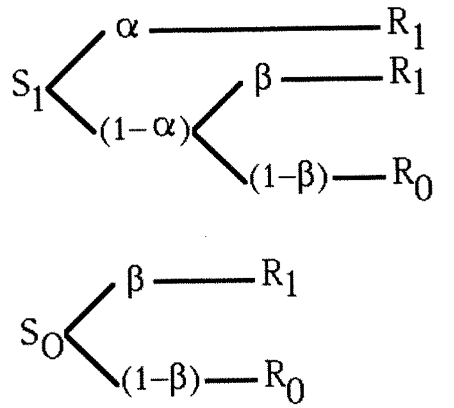

Figure 1. The Fechnerian high-threshold model (Blackwell, 1963) in which a signal (S_1) may be detected (with probability α) and evoke a yes (R_1) response, or the signal may not be detected and the subject guesses yes (with probability β) or no (R_0). Nonsignals (S_0) always lead to nondetections and similar guessing. Note that when a signal is detected the response is made with certainty, because only S_1 can produce a detection.

indicates the set of performances predicted for a particular value of the *sensitivity parameter* α as the *guessing parameter* β goes from 0 to 1. Also shown in Figure 2 are illustrative curvilinear OCs from the basic equal variance SDT model of Green and Swets (1966). This model also has two parameters: a sensitivity parameter d' and a *response criterion parameter* c, with

$$P(R_1 \mid S_1) = \Phi(c + d') \tag{4}$$

and

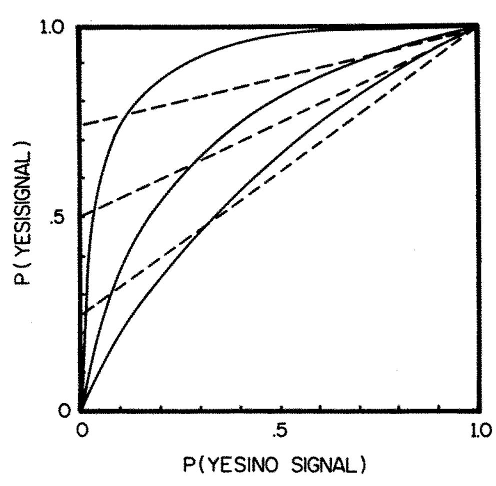

Figure 2. Illustrative linear operating characteristics (OCs) from the high-threshold (HT) model and curvilinear OCs from the Green and Swets (1966) equal variance statistical decision theory (SDT) model. The probability (P) of actually detecting the signal (α) under the HT model is the intercept of the linear OC where $P(\text{yes} \mid \text{no signal}) = 0$. Thus, if the data conform to the SDT model and one tries to fit it with the HT model, subjects who are instructed to be progressively more conservative will yield progressively lower estimates of α.

$$P(R_1 \mid S_0) = \Phi(c), \tag{5}$$

where Φ is the standard normal cumulative function; that is, $\Phi(c)$ is the area below c under a standard normal (Z) distribution. Note that the parameter c is defined here such that instructing a subject to be more liberal in reporting a signal would increase the value of c. Each OC for this model indicates the possible performances for a particular value of d' as c goes from minus to plus infinity.

Note that with the SDT model both hits and false-alarms approach zero as the decision criterion c goes to minus infinity, whereas in the HT model they approach the point (α, 0) as the guessing parameter β approaches 0. Furthermore, each SDT OC intersects every HT OC. If instructions to be more or less conservative in reporting a signal actually drive performance along an SDT OC, estimates of both sensitivity (α) and bias (β) that are based on the HT model would go from 0 to 1 as the subject went from being highly conservative to highly liberal, whereas the SDT model would yield a fixed estimate of sensitivity (d') and only the criterion parameter (c) would vary. Thus, the two models provide very different interpretations of the same data. This is why it is important to consider the extensive body of data indicating that the SDT model provides a more valid representation of both auditory detection and recognition memory data than does the HT model.

Rejecting the HT Model as a Model of Signal Detection

Green and Swets (1966) cited three principle types of data that are inconsistent with the HT model and support the SDT model. First, instructing a subject to be more or less liberal in reporting a signal does not move their performance along a linear OC as in Figure 2 but rather a curvilinear OC like those in the same figure. Second, estimates of α that are based on yes–no data and on n-interval, forced-choice data are not equivalent, whereas estimates of d' that are based on the two paradigms are equivalent. (In an n-interval, forced-choice task, the subject must decide which one of n intervals contained a signal.) Third, if a subject's first choice in an n-interval, forced-choice task ($n \geq 3$) is wrong, a second choice will be correct more often than chance. This is inconsistent with the HT model because an error on the first choice implies that the signal was not detected (all n intervals evoked a nondetection), making the second choice a pure guess. In contrast, the SDT model predicts above chance accuracy levels for second choices.

Batchelder and Riefer (1990) did not acknowledge these specific shortcomings of the HT model, although they referenced Laming (1973), who does. More important, however, they failed to reference any of the studies that have shown the HT model to be a poor model of recognition memory.

Rejecting the HT Model as a Model of Recognition Memory

Treating old items as signals (S_1) and new items as nonsignals (S_0) allows one to apply both the HT and SDT models to recognition memory data. Again the HT model is clearly inferior to the SDT model for reasons similar to those in psychophysical studies. First, many studies have shown that inducing a subject to be more or less liberal in reporting an item as old produces shifts in performance best described by the curvilinear OCs of SDT rather than the linear OCs of the HT model. This is illustrated in Figure 3, which presents data from a recent study by Ratcliff, Sheu, and Gronlund (1992). These results are typical of the data obtained in earlier recognition memory studies (e.g., see Egan, 1958; Glanzer & Adams, 1990; Lockhart & Murdock, 1970; Mandler & Boek, 1974; Murdock, 1974, 1982; Murdock & Dufty, 1972; Shiffrin, Ratcliff, & Clark, 1990). As shall be seen, its linear OC is the most serious shortcoming of the HT model, because it leads to an inevitable confounding of sensitivity and bias measures. Second, estimates of α that are based on yes–no data and on n-interval, forced-choice data are not the same, whereas estimates of d' normally are (Green & Moses, 1966; Yonelinas, Hockley, & Murdock, 1992). Third, if a subject's first choice among n alternative items as the single old item is wrong, a second choice is well above chance as predicted by the SDT model but not by the HT model (Brown, 1965). Thus, there is considerable evidence that the SDT model is a more valid characterization of recognition memory than the HT model.

The Multinomial as an HT Recognition Memory Model

Batchelder and Riefer (1990) characterized the most general form of the multinomial model with three tree diagrams like those in Figure 4. Here S_1, S_2, and S_N denote, respectively, an old source one, old source two, and new test item, and R_1, R_2, R_N are the three corresponding identification responses in a source-monitoring task. There are a total of seven theoretical parameters—D_1, D_2, d_1, d_2, a, b, and g—indicating the probabilities of specific paths through the trees. The D_1 and D_2 parameters are said to represent a subject's probabilities of detecting as old an S_1 or S_2 item, respectively. Similarly, d_1 and d_2 represent the probabilities of discriminating the source of an S_1 or S_2 item. The remaining parameters a, b, and g are termed *nuisance parameters,* denoting various response biases.

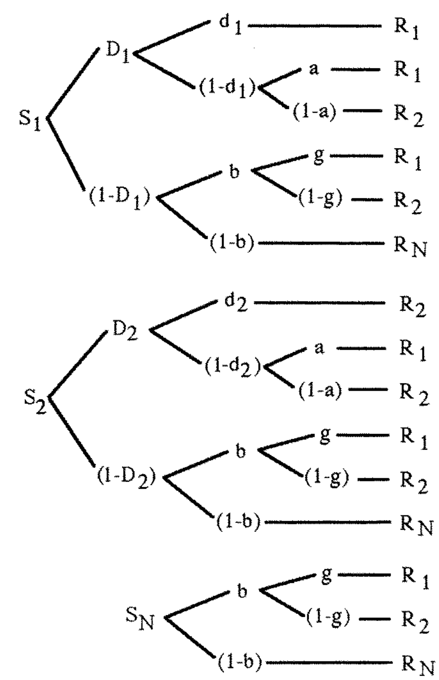

Figure 4. Tree diagrams defining the multinomial model. S_1 = old source one test item; S_2 = old source two test item; S_N = new test item; D_1, D_2, d_1, d_2, a, b, and g are theoretical parameters; R_1, R_2, and R_N are the identification responses corresponding to S_1, S_2, and S_N, respectively, in a source-monitoring task.

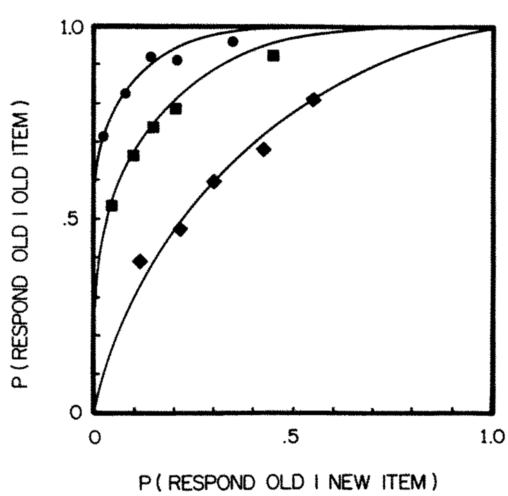

Figure 3. Representative data from a recognition memory study by Ratcliff, Sheu, and Gronlund (1992). The three statistical decision theory operating characteristics (OCs) reflect differences in the number of times an item was presented before the recognition test: five times (circular points), one time (square points), and one time intermixed with repeated items (diamond points), which further reduces recognition. The five data points defining each curve reflect variation in the proportion (P) of old (versus new) items in the test series. The higher that proportion, the more liberal the subjects; performances moved along the OC from left to right.

A major component of the Batchelder and Riefer (1990) article is the development of statistical techniques for estimating theoretical parameters, including ways of simplifying the model if estimates of some parameters are not significantly different (e.g., if estimates of d_1 and d_2 do not appear to differ, they can be replaced by a single parameter d). These are very interesting and sophisticated statistical techniques, and if the multinomial model was a valid model of source monitoring, these estimates would be valid cognitive measures.

What Batchelder and Riefer (1990) failed to make clear is that their multinomial model characterizes recognition (discriminating old from new items) in exactly the same way as the discredited Fechnerian HT model. This becomes clear if one simply ignores the specific source judgment and combines R_1 and R_2 responses (because both imply that the subject believes the item is old). The trees in Figure 4 then reduce to those shown in Figure 5, which is the same stochastic process shown in Figure 1 with D_i corresponding to α and b to β. Thus, regardless of the values of d_1, d_2, a, and g, the multinomial model predicts the same linear OCs for recognition as those in Figure 2. Specifically, just as Equations 1 and 2 lead to Equation 3,

$$P(R_1 \cup R_2 | S_i) = D_i + (1 - D_i) \cdot P(R_1 \cup R_2 | S_N), \quad (6)$$

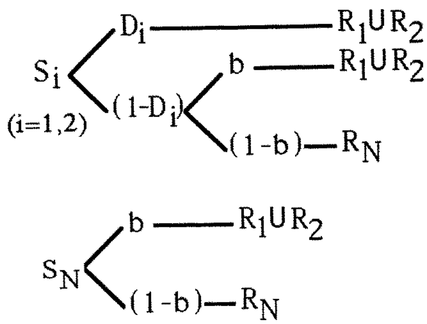

Figure 5. If one ignores source judgments by combining R_1 and R_2 responses, the tree diagrams in Figure 4 reduce to those shown here. Note that this is simply the high-threshold model shown in Figure 1 with D_i replacing α and b replacing β. S_1 = old source one test item; S_2 = old source two test item; S_N = new test item; D_1, D_2, and b are theoretical parameters; R_1, R_2, and R_N are the identification responses corresponding to S_1, S_2, and S_N, respectively, in a source-monitoring task.

for i = 1 or 2. Thus, the HT OC for recognition memory is a straight line from the point (1, 1) to (D_i, 0); that is, $P(R_1 \cup R_2 | S_i) = D_i$ when $P(R_1 \cup R_2 | S_0) = 0$. ($D_1$ may differ from D_2 defining two linear OCs, one for S_1 items and the other for S_2 items, just as two intensities of an auditory signal could yield two OCs and two d' values, using the SDT model.)

As cited earlier, there are many recognition memory studies that refute the HT model's prediction of linear recognition OCs in favor of the curvilinear OCs of the SDT model (Figure 2), and most of these data were published before the Batchelder and Riefer (1990) article. With no reference to this work, Batchelder and Riefer tested their model by fitting it to several sets of source-monitoring data and then giving a psychologically plausible interpretation of each set of data that was based on the theoretical parameter estimates. This does not seem to be a particularly demanding test of the model. The fact that parameters can be estimated from the data does not mean that they are valid measures, even if they provide the basis for a plausible interpretation.

The issue of plausability versus validity can be illustrated in terms of the two data points shown in Figure 6. They are part of a reality-monitoring study reported by Harvey (1985) and evaluated by Batchelder and Riefer (1990). Subjects were asked to either speak (S_1), or imagine speaking (S_2), a set of words; later they were shown these old words intermixed with new ones (S_N). The two source-monitoring performances (data points) in Figure 6 indicate the tendencies of manic patients to report S_2 items as old (R_1 or R_2) or new (R_N). The open circle represents thought-disordered patients, and the solid circle represents non–thought-disordered patients. According to the estimates of D_2 and b that are based on the multinomial HT model, the two groups differed in both sensitivity (D_2) and bias (b); for example, the non–thought-disordered manic patients were about twice as likely to detect an old S_2 item as were the thought-disordered ones (Batchelder & Riefer's conclusion). However, estimating d' and c of the SDT model that are based on the same data suggests a very different interpretation: Both groups had the same sensitivity (d'), but the thought-disordered patients were simply more conservative in reporting an item as old. Although the interpretations suggested by both models could be considered plausible, the results of the many recognition memory studies cited earlier clearly show that the curvilinear OCs of the SDT model are more valid (consistent with recognition data). Thus, fitting such data with the linear OCs of the HT model confounds the effects of biasing instructions with factors that actually alter a subject's memory for old items. Specifically, as can be seen in Figures 2, 3, and 6, simply instructing a subject to be more conservative in reporting items as old will lead to lower estimates of D_i, the probability of actually detecting an old item is old rather than having to guess whether it is or is not (see Figure 4).

Batchelder and Riefer (1990) asserted that the multinomial model has "certain advantages over traditional methods for measuring source memory because empirical statistics are often confounded by different cognitive processes and thus are not a pure measure of any specific one" (p. 560). Indeed, measures (empirical statistics) have been used in some studies of source monitoring whose computation involves some of the same numbers, such that the measures are necessarily correlated. However, the identification and measurement of different cognitive processes involves much more than this. Identifying a dissociation between the effects of two experimental variables is one basis for asserting that they influence different cognitive processes (see Ashby & Townsend, 1986, or Kadlec & Townsend, 1992, for a discussion of this issue). For example,

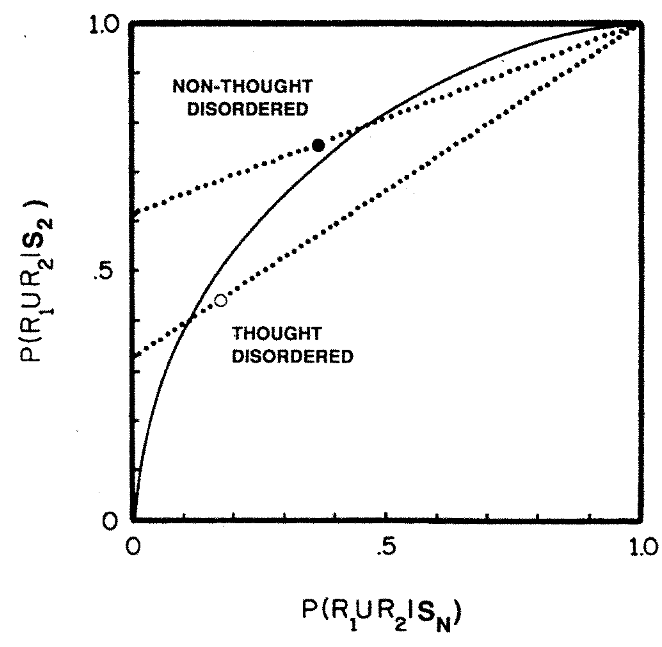

Figure 6. Harvey (1985) showed thought disordered and nonthought disordered manic patients' words that they had spoken earlier (S_1) or imagined speaking (S_2), mixed in with new words (S_N) and asked them to identify each word by responding R_1, R_2, or R_N, respectively. The graph shows how the two groups of patients differed in their relative tendencies to label S_2 and S_N words as old (respond R_1 or R_2). According to the high-threshold model, the groups differed in both sensitivity (D_2) and bias (b), whereas they differed only in bias (c) according to the statistical decision theory model. P = probability.

conjoint measurement techniques (Luce & Tukey 1964; or see Michell, 1990) are designed to identify such dissociations. The Green and Swets (1966), equal variance, SDT model is a form of stochastic conjoint measurement because it indicates that a certain transformation of the dependent variable in detection studies reveals an underlying additivity in the effects of certain independent variables. Specifically, signal variation (signal present or absent) and biasing instructions (liberal or conservative) will have an additive effect on performance if the proportion of detection (yes) responses, p, is transformed to Z_p such that

$$p = \Phi(Z_p), \tag{7}$$

with $\Phi(Z_p)$ the standard normal, cumulative density function. The resulting additivity is illustrated graphically in Figure 7. Inducing a subject to be more or less liberal in reporting signals influences the bias measure (c) but not the sensitivity measure (d') and vice versa. This double dissociation of effects suggests that signal variation and biasing instructions influence different underlying cognitive processes. Similar arguments can be made in recognition memory studies where instructions to be more or less liberal in deciding that an item is old have little effect on d', whereas how long old items were originally studied does (e.g., Shiffrin et al., 1990). This is not to argue that a dissociation of factors necessarily implies their underlying additivity, simply that the SDT analysis often demonstrates both.

Is There an SDT Model for Source Monitoring?

Batchelder and Riefer (1990) acknowledged that source monitoring is similar to detection tasks in which the signal may be at either of two frequencies (often called a *recognition task*) and that there are several ways to develop models for such tasks. However, in the interest of simplicity they decided to build their theory of source monitoring on the HT model. This tends to obscure the fact that the decision problem in source monitoring is fundamentally more complex than that represented in the basic Green and Swets (1966) SDT model of signal detection. In

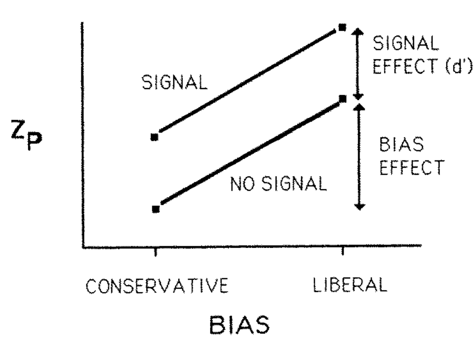

Figure 7. According to the statistical decision theory model, transforming the dependent variable p (the proportion of yes responses) to Z_p (Equation 7) reveals an additive effect of the two independent variables (signal variation and bias instructions), with d' (the difference between the transformed hit and false-alarm rates) the effect of presenting a signal rather than no signal. More generally, there is a double dissociation between the effect of biasing instructions and signal variation on the dependent measures d' and c.

that model, a subject's task is represented as a two-hypotheses statistical decision problem: Was there a signal (H_1), or not (H_0), given some sensory sample? All the information in an n-dimensional sample, X, relevant to deciding between two hypotheses, H_1 and H_0, can be reduced to a single number, the *likelihood ratio,* $1(X)$, where

$$1(X) = \frac{P(H_1|X)}{P(H_0|X)} \tag{8}$$

(Neyman & Pearson, 1933). A decision favoring H_1 is made only if $1(X)$ exceeds some critical value, termed the *decision criterion.* The decision parameter c used earlier in Equations 4 and 5 corresponds to a decision criterion except that it is defined here such that it increases as the subject becomes more liberal.

In contrast, deciding whether or not an auditory signal, at one of two possible frequencies, has been presented produces a three-hypotheses statistical decision problem: Was there a high-frequency signal (H_1), a low-frequency signal (H_2), or no signal (H_0), or analogously in source monitoring, is the item an old source one (H_1), old source two (H_2), or new (H_N) item. The information in a sensory sample (or memory sample) relevant to this type of decision cannot be reduced to a single likelihood ratio. Instead, at least two numbers are required, for example,

$$1_1(X) = \frac{P(H_1|X)}{P(H_0|X)} \tag{9}$$

and

$$1_2(X) = \frac{P(H_2|X)}{P(H_0|X)}. \tag{10}$$

This means that each sensory sample X could be thought of as a point in a two-dimensional space with coordinates $1_1(X)$ and $1_2(X)$. There are three, joint density, sampling distributions of these points, one under each hypothesis. Instead of the decision criterion being a point (c), it is defined by lines separating the two-dimensional sampling space into three response regions. Alternatively, subjects have been viewed as evaluating samples from two "channels," one "tuned" to S_1 and the other tuned to S_2. In either case, the sampling space is two-dimensional and the response regions are defined by lines rather than a point (Green & Birdsall, 1978; Tanner, 1956).

A number of SDT models for detecting and discriminating two-dimensional stimulus patterns are reviewed in Macmillan and Creelman (1991, chap. 10). However, the most extensive development of such multidimensional models can be found in Ashby (1992), particularly articles by Kadlec and Townsend (1992) and Thomas and Olzak (1992). These articles explicitly considered the two-frequency auditory detection task (e.g., did a tone signal occur, and, if so, was it the high or low frequency?). This is the task Batchelder and Riefer (1990) saw as being analogous to source monitoring (e.g., is this an old item, and, if so, is it from Source A or Source B?). Thus, the experimental and analytical methods developed within the multidimensional SDT framework may provide theoretical measures that more accurately reflect the underlying complexity of source monitoring than does the HT multinomial model.

Conclusions

Batchelder and Riefer (1990) pointed out that none of the studies of source monitoring they surveyed used a substantive model to analyze its data. In general, an explicit mathematical model is to be preferred to a vague, verbal one, and theirs is apparently the first quantitative model applied to source monitoring. However, an ill-fitting (invalid) mathematical model may lend a spurious precision to the interpretation of the experimental phenomena. Of course, it is sometimes useful to use a simpler and more mathematically tractable model rather than one that is less tractable but more consistent with the data. After all, all models are only useful approximations to the empirical phenomena under study. Nevertheless, there are different degrees of approximation, and if a simplified version of a model differs in a clear and systematic fashion from the data, as does the HT recognition memory model, this fact should be made clear. Batchelder and Riefer's (1990) minimal explanation of their model's relation to the Fechnerian HT model, or of the experimental evidence against it as a model of recognition, may encourage readers to place more confidence in the psychological validity of their measures than is appropriate. Furthermore, appending a rather complicated (four parameter) source discrimination process onto a demonstratively poor model of recognition memory seems unlikely to yield valid measures of source discrimination. I suggest that models that are based on multidimensional SDT may yield such measures.

References

Ashby, F. G. (1992). *Multi-dimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564.

Blackwell, H. R. (1963). Neural theories of simple visual discrimination. *Journal of the Optical Society of America, 53,* 129–160.

Brown, J. (1965). A comparison of recognition and recall by a multiple-response method. *Journal of Verbal Learning and Verbal Behavior, 4,* 401–408.

Egan, J. P. (1958). *Recognition memory and the operating characteristic.* (Tech. Rep. No. AFCRC-TN-58-51). Bloomington: Indiana University Hearing and Communication Laboratory.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Green, D. M., & Birdsall, T. G. (1978). Detection and recognition. *Psychological Review, 85,* 192–206.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66,* 228–234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Kreiger.

Harvey, P. D. (1985). Reality monitoring in mania and schizophrenia. *The Journal of Nervous and Mental Disease, 173,* 67–72.

Kadlec, H., & Townsend, J. T. (1992). Signal detection analyses of dimensional interactions. In F. Gregory Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–227). Hillsdale, NJ: Erlbaum.

Laming, D. (1973). *Mathematical psychology.* San Diego, CA: Academic Press.

Lockhart, R. S., & Murdock, B. B., Jr. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74,* 100–109.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1,* 1–27.

Macmillan, N., & Creelman, C. (1991). *Detection theory: A users guide.* Cambridge, England: Cambridge University Press.

Mandler, G., & Boek, W. J. (1974). Retrieval processes in recognition. *Memory & Cognition, 2,* 613–615.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Erlbaum.

Murdock, B. B. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B., Jr., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology, 94,* 284–290.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A, 231,* 289.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Tanner, W. P., Jr. (1956). Theory of detection. *The Journal of the Acoustical Society of America, 28,* 882–888.

Thomas, J. P., & Olzak, L. A. (1992). Simultaneous detection and identification. In F. Gregory Ashby (Ed.), *Multi-dimensional models of perception and cognition.* Hillsdale, NJ: Erlbaum.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 345–355.

Received August 21, 1992
Revision received April 2, 1993
Accepted April 29, 1993 ■