

# Response time distributions in multidimensional perceptual categorization

W. TODD MADDOX

*Arizona State University, Tempe, Arizona*

F. GREGORY ASHBY

*University of California, Santa Barbara, California*

and

LAWRENCE R. GOTTLÖB

*Arizona State University, Tempe, Arizona*

Three speeded categorization experiments were conducted using separable dimension stimuli. The form of the category boundary was manipulated across experiments, and the distance from category exemplars to the category boundary was manipulated within each experiment. Observers completed several sessions in each experiment, yielding 300–400 repetitions of each stimulus. The large sample sizes permitted accurate estimates of the response time (RT) distributions and RT hazard functions. Analyses of these data indicated: (1) RT was faster for stimuli farther from the category boundary, and this stochastic dominance held at the level of the RT distributions; (2) RT was invariant for all stimuli the same distance from the category boundary; (3) when task difficulty was high, errors were slower than correct responses, whereas this difference disappeared when difficulty was low; (4) small, consistent response biases appeared to have a large effect on the relation between correct and error RT; (5) the shape of the RT hazard function was qualitatively affected by distance to the category boundary. These data establish a rich set of empirical constraints for testing developmental models of categorization RT.

Fast, accurate categorization is fundamental to survival (Ashby & Maddox, 1998). Whenever we define an object as a “kind” of thing, we are categorizing. In keeping with its important role in perception and cognition, several powerful theories have been proposed and model-based instantiations developed to predict categorization performance. These include, among others, prototype (Anderson, 1991; Homa, Dunbar, & Nohre, 1991; Reed, 1972), exemplar (see Estes, 1994, and Nosofsky, 1992, for reviews), and decision bound models (e.g., Ashby, 1992a; Ashby & Maddox, 1993, 1998; Maddox, 1995; Maddox & Ashby, 1993). Nearly all these models focus exclusively on categorization accuracy as the dependent variable.

Although categorization accuracy provides important information about the process of categorization, the observer’s response time (RT) often provides a richer source

of information. For example, an observer might respond with the same accuracy level for two category exemplars, but with different RT distributions (e.g., Laming, 1968; Luce, 1986; Welford, 1968). To date, few rigorous theories of categorization RT exist; however, some are currently being developed, and initial tests are being conducted (e.g., Ashby & Maddox, 1994; Maddox & Ashby, 1996; Nosofsky & Palmeri, 1997). The goal of this article is to provide a rich base of categorization RT data that can be used to guide the development and testing of new and emerging models of categorization RT. Specifically, the aim is to identify a set of empirical constraints that must be predicted by any viable theory of categorization RT.

A robust empirical finding is that correct-response mean RT tends to decrease as the distance between the exemplar and the category boundary increases (Bornstein & Monroe, 1980; Cartwright, 1941; see also Ashby, Boynton, & Lee, 1994). That is, exemplars that are far from the category boundary yield (on average) fast categorization responses, and exemplars near the category boundary yield (on average) slow categorization responses. Ashby and Maddox (1991, 1994) formalized this notion and called it the *RT–distance hypothesis*. This is an important empirical finding, but it has been tested only on a fairly weak statistic of the data—namely, mean RT. Higher order statistics, such as the RT distribution and RT hazard function, provide a richer source of information about categorization RT and thus yield more powerful empirical constraints on

---

This research was supported in part by an Arizona State University Faculty-Grant-in-Aid and by National Science Foundation Grants DBS92-09411 and SBR-9514331. The authors would like to thank Donald Laming, Neil Macmillan, and several anonymous reviewers for helpful comments on an earlier version of this manuscript, J. D. Balakrishnan and Thomas Fikes for helpful comments and suggestions about the hazard function routines, and Corey Bohil for help with some of the figures. L. R. Gottlob is now at Duke University Medical Center, Center for the Study of Aging and Human Development, P.O. Box 3003, Durham, NC 27710. Correspondence concerning this article should be addressed to W. T. Maddox, Department of Psychology, University of Texas, 330 Mezes Hall, Austin, TX 78712 (e-mail: maddox@psy.utexas.edu).

categorization RT theories (Ashby, Tein, & Balakrishnan, 1993; Townsend, 1991; Townsend & Ashby, 1978, 1983). Lower order statistics, such as mean RT, are useful because accurate estimates require (relatively) little data, and procedures for testing statistical significance (such as analysis of variance [ANOVA]) are well developed. Most distributional level statistics require much larger sample sizes, and some, such as the RT hazard function, do not have well-developed procedures for testing statistical significance. Even so, relations among lower level statistics are implied by relations among higher level statistics, whereas the reverse is not true. For example, an ordering of two cumulative distributions implies an ordering of the means, but an ordering of the means implies nothing about the cumulative distributions (e.g., Townsend, 1991). Thus, it is advantageous to examine both higher and lower level statistics whenever possible.

Three experiments are reported in this article. In all three, the stimuli were circles of varying size with a radial line of varying orientation. In each experiment, there were two categories of nine exemplars; each and every stimulus was presented to each observer 300–400 times. These large sample sizes made it possible to estimate the RT distributions for individual stimuli and individual observers.<sup>1</sup> Our analysis of the resulting data focused on four different empirical issues.

First, we examined the relationship between RT and distance to the category boundary. There is good evidence that mean RT decreases with distance to boundary (i.e., the RT–distance hypothesis), but it is unknown whether this relationship extends to higher level distributional statistics. Another goal of this analysis was to examine a corollary of the RT–distance hypothesis—namely, that RT is invariant for all stimuli the same distance from the boundary. Currently, it is unknown whether this prediction holds even at the level of mean RT (although see Ashby et al., 1994).

Second, we examined the relation between correct-response and error RT. Many categorization studies (although by no means all) utilize highly discriminable categories that yield small error rates. In these situations, it is impossible to estimate mean error RT. A complete theory of categorization RT must make predictions about correct and error RT and about the relation between the two (e.g., are categorization errors always slower than correct responses?). Currently, there are no data that provide estimates of correct and error mean RTs to the same stimuli in a multidimensional categorization experiment.

Third, we examined the effect of category response bias on RT. Small but consistent response biases are often observed in categorization studies (e.g., Huttenlocher, Hedges, & Duncan, 1991). In fact, most models of categorization accuracy include a parameter to account for response bias (e.g., Ashby, 1992a; Nosofsky, 1986). Even so, little effort has been made to examine the influence of category response bias on categorization RT.

Finally, we examined how the nature of the category boundary affects categorization RT. It is known that the nature of the category boundary (e.g., whether the boundary

is linear or quadratic) affects category learning and asymptotic accuracy (Ashby & Maddox, 1990, 1992; Maddox & Ashby, 1993, 1996), but virtually nothing is known about effects on RT. We examined three qualitatively different types of category boundaries. In the selective attention experiment, the category boundary was linear but was positioned in such a way that the correct strategy was to attend selectively to one stimulus dimension and ignore the other. In the linear integration experiment, the category boundary was again linear but was positioned in such a way that the correct strategy was to attend (approximately) equally to each stimulus dimension. In the nonlinear integration experiment, the category boundary was highly nonlinear. As we will see, the complexity of the category boundary has a large (and systematic) effect on the RT distributions.

To summarize, the goal of the present research was to provide a detailed examination of RT distributions in multidimensional perceptual categorization. In particular, we were interested in the effects of several factors on lower and higher order properties of the RT distributions. Our aim was to provide a rich database of empirical constraints that must be predicted by any viable theory of categorization RT.

The next section provides an overview of several important properties of RT distributions and details the stochastic dominance relations used to investigate the relation between RT and distance to the category boundary. The third section is devoted to the experimental method, and the fourth section is devoted to the results. Finally, we close with some general comments on the implications of our findings for current and future theories of categorization RT.

## RESPONSE TIME DISTRIBUTIONS AND TESTS OF STOCHASTIC DOMINANCE

Every presentation of the same category exemplar leads to a unique RT.<sup>2</sup> The resulting data can be described at many different levels. This article focuses on four. The RT density function, denoted by  $f(t)$ , gives the likelihood that RT equals  $t$ , for each specific value of  $t$ . The cumulative RT distribution at time  $t$ , denoted by  $F(t)$ , gives the probability that the observed RT is less than or equal to  $t$  [i.e.,  $F(t) = P(\text{RT} \leq t)$ ]. The hazard function, denoted by  $h(t)$ , defines the probability that the response will occur in the next instant given that it has not yet occurred. More formally, the hazard function is defined as  $h(t) = f(t)/[1 - F(t)]$ . Finally, the mean or expected RT defines the average RT and is denoted by  $E(\text{RT})$ .

One of the major goals of this article is to examine whether stimuli that are close to the category boundary yield RTs that are stochastically greater than RTs for stimuli that are farther from the boundary. Theoretically, there are several levels at which this stochastic dominance can be tested (Ashby et al., 1993; Townsend, 1991; Townsend & Ashby, 1978, 1983). One of the weakest is at the level of the mean or expected RT. Let  $E(\text{RT}_i)$  represent the mean RT for Stimulus  $i$ , and let  $d_{iB}$  represent the distance between Exemplar  $i$  and the category boundary. We can then

conclude that the RT–distance hypothesis is supported at the mean RT level if, for all  $d_{iB} < d_{jB}$ ,

$$E(RT_j) \leq E(RT_i). \quad (1)$$

A stronger form of stochastic dominance is an ordering at the level of the cumulative RT distribution functions. The RT–distance hypothesis is supported at the cumulative distribution function level if, for all  $d_{iB} < d_{jB}$ ,

$$F_j(t) \geq F_i(t), \text{ for all } t > 0. \quad (2)$$

An ordering of the cumulative RT distributions is a stronger RT–distance effect than an ordering of the mean RTs because the former implies the latter—that is, an ordering of the cumulatives implies an ordering of the means—but the latter does not imply the former (Townsend, 1991; Townsend & Ashby, 1978, 1983). An even stronger form of stochastic dominance holds if the RT hazard functions are ordered by distance to bound—that is, if, for all  $t > 0$ , and all  $d_{iB} < d_{jB}$ ,

$$h_j(t) \geq h_i(t). \quad (3)$$

An ordering of the hazard functions implies an ordering of the cumulative distributions (Equation 2), but an ordering of the cumulatives does not imply an ordering of the hazard functions. Not only is a hazard function ordering informative, but the shape of the hazard function also provides important information about the nature of perceptual processing (e.g., Ashby et al., 1993; Luce, 1986). An even stronger form of stochastic dominance is at the level of the likelihood ratio

$$l(t) = f_i(t)/f_j(t). \quad (4)$$

The RT–distance hypothesis is supported at the likelihood ratio level if, for all  $d_{iB} < d_{jB}$ ,

$$l(t) \text{ is nondecreasing in } t. \quad (5)$$

This is the strongest form of stochastic dominance considered in this article because if Equation 5 holds, it implies the other three forms of stochastic dominance (i.e., Equations 1, 2, and 3). To summarize the stochastic dominance relations, when  $l(t)$  is nondecreasing for all  $t > 0$ , and  $d_{iB} < d_{jB}$ , then the following three relations are implied:

$$h_j(t) \geq h_i(t), \text{ for all } t > 0$$

$$F_j(t) \geq F_i(t), \text{ for all } t > 0$$

$$E(RT_j) \leq E(RT_i).$$

Each type of stochastic dominance relation was tested in data from three multidimensional perceptual categorization experiments.

## EXPERIMENTS

The stimuli used in all experiments were circles of varying diameter that contained a radial line of varying orientation (see Figure 1a). These stimulus components are thought to be separable (e.g., Garner, 1974; Garner & Fel-

foldy, 1970; Shepard, 1964; however, see Ashby & Lee, 1991; Ashby & Maddox, 1990). The stimulus ensemble consisted of 18 circular stimuli; it is displayed, along with a numbering scheme, in Figure 1b. In the selective attention experiment, the experimenter-defined categorization rule required the observers to ignore the orientation of the radial line and base their categorization judgment solely on the diameter of the circle (see Figure 2a). In the linear integration experiment, the categorization rule required the observers to attend to both stimulus components and use a linear decision bound (see Figure 2b). In the nonlinear integration experiment, the categorization rule required attention to both dimensions and the use of a nonlinear decision bound (see Figure 2c).

During presentation of the data analyses, the selective attention and linear integration experiments will be discussed in parallel, whereas the nonlinear integration experiment will be treated separately. This approach is taken because the observers, stimuli, and experimental procedures were identical in the selective attention and linear integration experiments. The only difference between the selective attention and linear integration experiments was in the stimulus-to-category mappings.

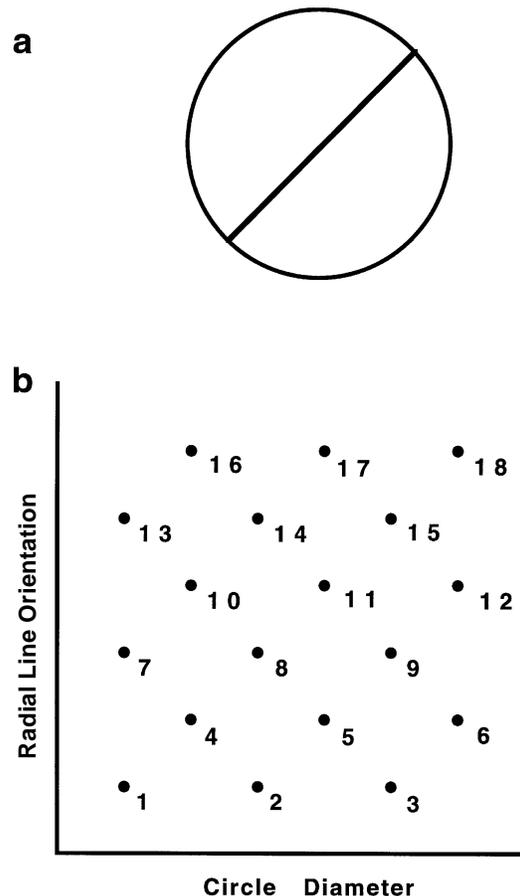


Figure 1. (a) Sample stimulus. (b) Stimulus structure and numbering scheme.

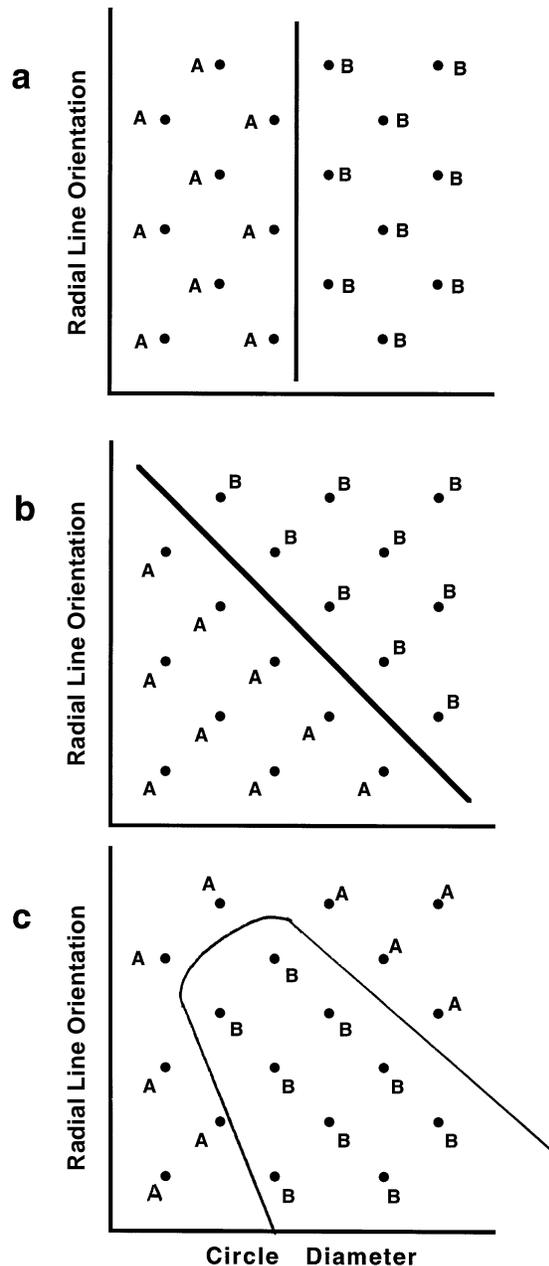


Figure 2. Stimulus-response mappings and experimenter-defined category boundary for the (a) selective attention experiment, (b) linear integration experiment, and (c) nonlinear integration experiment.

## GENERAL METHOD

### Observers

All observers in the experiments were solicited from the Arizona State University community. The observers were paid \$5 for each experimental session. All observers had 20/20 vision or vision corrected to 20/20. The same 4 observers participated in the selective attention and linear integration experiments. All observers completed the selective attention experiment first. Three observers participated in the nonlinear integration experiment. Observers 1 and 2

from the selective attention and linear integration experiments participated as Observers 1 and 2, respectively, in the nonlinear integration experiment. Observer 3 was the first author. The number of trials per session and the number of sessions completed by each observer are detailed in Table 1. In each experiment, the first few sessions were considered practice and were excluded from the subsequent analyses.<sup>3</sup> In the selective attention experiment, the first session was considered practice. In the linear integration experiment, the first three or four sessions (depending on the observer's performance) were considered practice. In the nonlinear integration experiment, the first five or six sessions (depending on the observer's performance) were considered practice.

### Stimuli

The stimulus ensemble consisted of 18 circles, each with an embedded radial line. These stimuli represent a subset taken from 36 stimuli that were constructed from the factorial combination of six levels of circle diameter with six levels of radial line orientation. In the selective attention and linear integration experiments, the six diameters were 100, 109, 118, 127, 136, and 145 pixels, and the six orientations were .126, .201, .276, .352, .427, and .503 radians. In the nonlinear integration experiment, the six diameters were 100, 115, 130, 145, 160, and 175 pixels, and the six orientations were .126, .251, .377, .503, .628, and .754 radians. The 18 experimental stimuli were selected by taking 3 equally spaced stimuli from a given level of one stimulus component (see Figure 1b). The component level structure of the stimuli in the selective attention and linear integration experiments was such that three levels of distance to bound could be specified. We arbitrarily assigned the labels *small*, *medium*, and *large* to these distances. The stimuli classified into the three distance categories for the selective attention and linear integration experiments are outlined in Table 2. In all experiments, Category A contained 9 stimuli, and Category B contained 9 stimuli. The category assignments along with the category boundary are presented in Figure 2. The average visual angle was about 1°. The stimuli were computer generated and displayed on a noninterlaced Super VGA monitor in a dimly lit room.

### Procedure

At the start of each experimental session, the observer was shown the 18 stimuli along with the category boundary in a form similar to that displayed in Figure 2. On every trial of the experiment, the observer's task was to categorize the stimulus as an exemplar of Category A or Category B. The observers were instructed "to respond as quickly as possible without sacrificing accuracy." Each trial of the experiment proceeded as follows. First, 1 of the 18 stimuli was selected at random (each with equal probability). A fixation point (i.e., a plus sign [+]) appeared in the center of the screen for 500 msec. The stimulus was presented for 250 msec, followed by a pattern mask that remained on the screen until the observer responded. The pattern mask was included to ensure that the maximum amount of perceptual processing on each trial was constant. The observer responded by pressing the key marked "A" for Category A or the key marked "B" for Category B. Each response was followed by a 1,000-msec display of the correct category label, a 1,000-msec blank screen, and then initiation of the next trial. Every 25 trials, the observer was allowed to rest and was provided with the cumulative accuracy score and correct-response mean RT. Each observer, when ready to continue, pressed a button, and the next block of 25 trials was initiated.

## RESULTS

### Selective Attention and Linear Integration Experiments

A major variable of interest in this study was distance to the category boundary. When comparing stimuli of

**Table 1**  
**Number of Sessions Completed by Each Observer in Each Experiment**

Observer	Selective Attention Experiment (600 Trials/Session)	Linear Integration Experiment (600 Trials/Session)	Nonlinear Integration Experiment (400 Trials/Session)
1	11	16	26
2	15	15	25
3*	13	15	26
4	13	15	

\*Observer 3 in the nonlinear integration experiment, the first author, was different from Observer 3 in the selective attention and linear integration experiments.

varying distances, we can specify three types of comparisons. Comparisons between small-distance stimuli and medium-distance stimuli are called *one-step near* comparisons. We use this term because they involve stimuli that are near the bound but that differ by one arbitrary unit of distance. Comparisons between medium- and large-distance stimuli are called *one-step far* comparisons, because they involve stimuli that are far from the bound but that differ by one arbitrary unit of distance. Finally, comparisons between small- and large-distance stimuli are called *two-step* comparisons.

It is worth noting that this “arbitrary unit” of distance was smaller in the selective attention experiment than in the linear integration experiment. In the selective attention experiment, the unit of distance was a function only of the circle diameter. In the linear integration experiment, on the other hand, the distance was determined by the circle diameter and the line orientation. Statistically, the smaller units in the selective attention experiment made it more difficult to identify differences in RT as a function of distance because the distances were smaller.

### Accuracy Analysis

Tables 3 and 4 display the accuracy rates by stimulus and observer for the selective attention experiment and the linear integration experiment, respectively. Table 5 displays the selective attention and linear integration accuracy rates for each observer broken down by distance to bound and by category. The results can be summarized as follows. First, for each observer in both experiments, accuracy increased monotonically as distance to bound increased. Second, accuracy rates were similar in the selective attention and linear integration experiments, with a slight advantage for the selective attention experiment. Third, in some cases, the observers showed a bias for one

response over the other. This can be seen most clearly for Observers 2 and 3. For Observer 2, in both experiments, accuracy rates for Category B were higher than for Category A (selective attention experiment, advantage for Category B, 3.83%; linear integration experiment, advantage for Category B, 5.58%). This pattern held for each of the three distance-to-bound relations. Observer 3 showed a similar pattern; however, the bias was toward Category A instead of Category B (selective attention experiment, advantage for Category A, 5.01%; linear integration experiment, advantage for Category A, 2.41%). Observer 1 showed a moderate bias toward Category A, and Observer 4 showed a slight bias toward Category B, both in the linear integration experiment (Observer 1, advantage for Category A, 4.22%; Observer 4, advantage for Category B, 1.14%). As we will see shortly, these biases are mirrored nicely in the correct-response and error mean RT data, especially for Observers 2 and 3.

### Observer’s Category Boundary

Although a display of the stimulus configuration along with the experimenter-defined category boundary was provided for each observer, and accuracy rates were high, it is possible that the observers did not use the experimenter-defined category boundary. Because our classification scheme of small, medium, and large distances to the boundary is based on the experimenter-defined boundary, it is important to determine whether each observer used this boundary or at least one similar to the experimenter-defined boundary. To make this determination, we fit a linear and quadratic category boundary to each observer’s data. (The details of this procedure are outlined in the Appendix for the interested reader.) The experimenter-defined category boundary along with the “most parsimonious” boundaries (whether linear or quadratic) are displayed in Figure 3. If

**Table 2**  
**Stimuli Assigned to the Three Distance-to-Bound Categories**

Experiment	Category	Distance to Bound		
		Small	Medium	Large
Selective attention	A	2, 8, 14	4, 10, 16	1, 7, 13
	B	5, 11, 17	3, 9, 15	6, 12, 18
Linear integration	A	3, 5, 8, 10, 13	2, 4, 7	1
	B	6, 9, 11, 14, 16	12, 15, 17	18

Note—The stimulus numbering scheme is depicted in Figure 1b.

**Table 3**  
**Accuracy Rates (Acc) and Correct-Response (CR) and Error (E) Mean RTs by Stimulus and Observer for the Selective Attention Experiment**

Stimulus	Observer												Average		
	1			2			3			4			Acc	CR	E
	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E			
1	98.19	311	297	96.90	350	222	98.69	301	280	98.98	306	267	98.11	320	257
2	74.85	338	327	89.98	388	291	85.71	327	324	80.85	325	286	83.30	348	308
3	88.33	317	312	94.19	335	348	90.89	328	319	90.00	288	308	91.06	319	322
4	94.71	327	307	95.19	360	274	96.66	309	310	96.84	310	254	95.65	328	286
5	71.59	329	325	80.24	356	404	67.57	341	333	71.50	298	319	72.58	334	346
6	94.53	307	282	98.31	313	301	95.67	307	268	98.99	272	264	96.79	302	280
7	98.08	317	325	97.20	353	211	99.27	304	335	98.96	303	298	98.33	321	266
8	72.90	348	327	82.79	406	320	83.13	333	336	80.78	325	291	80.13	356	318
9	89.81	320	296	96.30	324	332	94.97	321	277	92.71	277	293	93.82	312	298
10	89.86	325	321	93.20	364	266	96.73	311	323	92.73	311	271	93.23	331	292
11	81.23	329	327	88.50	349	395	72.08	336	325	81.28	303	320	81.31	331	339
12	96.44	309	288	99.78	305	255	96.95	310	300	97.74	271	294	97.72	299	299
13	97.72	316	332	97.18	357	243	98.58	300	286	99.75	307	299	98.23	323	274
14	64.41	336	335	67.46	420	368	74.44	341	353	69.19	330	295	68.54	360	339
15	93.07	315	290	97.90	317	319	91.00	320	298	94.90	279	293	94.50	309	297
16	91.28	330	335	91.37	377	310	96.80	314	323	94.67	317	266	93.35	337	312
17	84.29	325	316	93.52	345	350	78.55	337	312	82.02	292	311	85.09	327	319
18	95.18	305	287	98.55	299	254	97.97	310	261	98.93	270	286	97.67	296	274
Average	87.48	322	321	92.18	348	338	89.79	318	323	89.94	298	299	89.95	324	321

Note—Correct-response and error mean RTs are in milliseconds.

the extra parameters of the quadratic category boundary did not provide a statistically significant improvement in fit over the linear category boundary, then the most parsimonious boundary was linear. However, if the extra parameters of the quadratic boundary did provide a statistically significant improvement in fit, then the most parsimonious boundary was quadratic. Figures 3a and 3b display boundaries for the selective attention and linear integration experiments, respectively. The dashed line in each

display represents the experimenter-defined category boundary. An examination of Figure 3 suggests that each observer's most parsimonious boundary differed only slightly from the experimenter-defined boundary. An examination of the effects of distance to the category boundary on RT requires only that the observer's category boundary preserved the distance-to-bound relations outlined above (i.e., small, medium, and large distances). Clearly, these distance relations were not violated for any

**Table 4**  
**Accuracy Rates (Acc) and Correct-Response (CR) and Error (E) Mean RTs by Stimulus and Observer for the Linear Integration Experiment**

Stimulus	Observer												Average		
	1			2			3			4			Acc	CR	E
	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E			
1	99.51	349	303	99.04	367	222	99.73	346	328	99.17	343	248	99.36	352	257
2	98.98	362	383	97.55	384	310	99.17	364	389	96.65	360	304	98.11	367	326
3	87.43	399	387	78.68	471	422	84.43	427	359	83.29	408	365	83.36	426	387
4	98.25	359	318	96.78	390	228	99.48	363	472	97.78	358	273	98.05	367	277
5	77.21	403	409	78.36	473	390	77.03	425	419	82.63	415	378	78.75	429	401
6	60.34	416	408	82.48	445	488	72.47	411	433	77.47	393	451	73.13	418	436
7	97.73	358	391	98.53	385	293	99.72	349	431	98.83	355	302	98.64	362	348
8	82.72	400	402	80.30	441	377	82.21	400	455	84.47	399	393	82.41	410	406
9	86.82	390	386	93.28	386	587	85.18	411	402	91.44	355	369	89.16	385	420
10	76.70	390	429	76.10	461	418	78.09	400	462	80.89	410	405	77.90	414	430
11	87.96	380	369	96.52	376	583	89.52	408	372	92.05	354	360	91.58	379	391
12	97.40	361	326	100.00	335	n.a.	97.21	374	349	99.25	319	291	98.51	346	331
13	86.82	383	431	85.54	447	379	93.30	381	435	81.41	394	392	86.87	401	404
14	82.42	391	379	92.62	393	480	77.36	433	378	87.37	367	394	85.48	394	397
15	97.35	350	347	100.00	334	n.a.	97.72	376	314	99.71	324	289	98.69	346	331
16	61.85	408	397	78.59	426	466	71.51	450	385	71.54	375	400	70.74	414	408
17	97.32	354	352	98.68	351	327	98.75	380	378	98.40	326	285	98.27	353	337
18	96.77	350	337	99.76	323	308	99.43	362	312	99.42	314	346	98.86	337	333
Average	87.42	375	398	90.86	394	425	89.13	389	409	90.08	362	390	89.36	380	405

Note—Correct-response and error mean RTs are in milliseconds.

**Table 5**  
**Accuracy Rates (Acc) and Correct-Response (CR) and Error (E) Mean RTs for the**  
**Selective Attention and Linear Integration Experiments by Observer, Distance to Bound, and Category**

Category	Selective Attention Experiment												Linear Integration Experiment											
	Distance to Bound												Distance to Bound											
	Small			Medium			Large			Average			Small			Medium			Large			Average		
	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E	Acc	CR	E
Observer 1																								
A	70.55	341	330	91.91	327	323	98.00	315	319	86.84	326	328	82.11	395	413	98.30	360	365	99.51	349	303	89.49	376	410
B	78.96	328	323	90.44	317	301	95.39	307	285	88.14	317	313	75.60	395	394	97.36	355	342	96.77	350	337	85.27	374	390
Average	74.78	334	327	91.20	322	311	96.72	311	296	87.48	322	321	78.91	395	403	97.83	357	351	98.20	349	332	87.42	375	399
Observer 2																								
A	79.91	403	342	93.29	367	287	97.09	353	226	90.25	372	317	79.90	458	399	97.63	386	268	99.04	367	222	88.03	420	389
B	87.28	350	392	96.15	325	338	98.86	306	278	94.08	326	373	88.82	403	496	99.58	340	327	99.76	323	308	93.61	371	492
Average	83.67	375	362	94.73	345	305	97.97	330	240	92.18	349	339	84.44	429	435	98.61	363	277	99.40	345	239	90.82	395	425
Observer 3																								
A	81.05	333	341	96.73	311	318	98.85	302	294	92.27	314	335	83.11	406	428	99.45	359	424	99.73	346	328	90.32	381	428
B	72.64	338	325	92.24	323	302	96.84	309	277	87.26	322	316	79.44	421	397	97.92	377	343	99.43	362	312	87.91	397	394
Average	76.87	336	331	94.53	317	307	97.85	305	281	89.79	318	324	81.32	413	412	98.68	368	360	99.59	354	317	89.13	389	409
Observer 4																								
A	76.84	327	291	94.70	313	266	99.23	305	284	90.11	314	287	82.56	405	387	97.74	358	293	99.17	343	248	89.50	380	379
B	78.30	298	317	92.61	281	299	98.55	271	285	89.77	282	311	83.95	368	406	99.10	323	287	99.42	314	346	90.64	345	402
Average	77.56	312	304	93.68	297	285	98.89	288	285	89.94	298	299	83.27	386	396	98.42	340	291	99.29	329	287	90.08	362	390

Note—Correct-response and error mean RTs are in milliseconds.

of the observers. A discussion of the best-fitting category boundaries, along with possible explanations for the moderate quadratic trend, is given in the Appendix.

### Stochastic Dominance Tests of the Response Time Data

One problem in testing RT hypotheses is that they make predictions about the relation of RT to the position of stimuli in the perceptual space, not the physical space. Since the perceptual space is unobservable, this could create problems in testing the RT hypothesis. Fortunately, much is known about the perceptual representation of the stimuli used in these experiments (at least about the mean perceptual effect). First, the Stevens exponent for length is very close to one, suggesting a close correspondence between the physical and perceived size of the stimulus. Second, psychological scaling (MDS) solutions for these stimuli, from many experiments, suggest a close correspondence between the physical and the perceptual space (Nosofsky, 1986; Nosofsky, Clark, & Shin, 1989; Shepard, 1964). Finally, since much of the present analyses require only an ordinal relation between RT and position of the stimulus in the stimulus space, this ordinal relation will remain intact under a wide range of monotonic transformations that might characterize the relation between the physical and the perceptual space (e.g., any strictly increasing monotonic transformation). Thus, these analyses do not require a close correspondence between the physical and perceptual spaces.

**Relation between correct-response and error RT.** Tables 3 and 4 show the correct-response and error mean RTs for each stimulus and observer for the selective attention and linear integration experiments, respectively.

Table 5 displays the correct-response and error mean RTs separately by distance to bound, experiment, and category. Working within standard univariate signal detection theory, Thomas (1971; see also Thomas & Myers, 1972) showed that the RT–distance hypothesis makes strong predictions about the relation between RT on correct and incorrect trials. Specifically, Thomas showed that, for most common perceptual distributions (e.g., normal), error RT will be longer than correct-response RT.<sup>4</sup> To test this hypothesis, we performed *t* tests on the correct-response and error mean RTs for each stimulus and observer. Averaged across observers and stimuli, 7% (selective attention experiment) and 15% (linear integration experiment) of the correct-response mean RTs were significantly faster than the error mean RTs (in support of the RT–distance hypothesis), and 29% (selective attention experiment) and 25% (linear integration experiment) of the correct-response mean RTs were significantly slower than the error mean RTs (providing evidence against the RT–distance hypothesis). Interestingly, the majority of the tests were not statistically significant and, thus, did not provide support for or evidence against the RT–distance hypothesis (64% in the selective attention experiment and 60% in the linear integration experiment). In the selective attention experiment, the correct-response and error mean RTs were very similar. In fact, averaged across stimuli and observers, the error mean RT was 3 msec faster than the correct-response mean RT. In the linear integration experiment, the error mean RTs were consistently longer than the correct-response mean RTs for all 4 observers. This difference was relatively large, averaging 25 msec across observers. The results from the linear integration experiment support the prediction of the RT–distance hypothesis that errors will

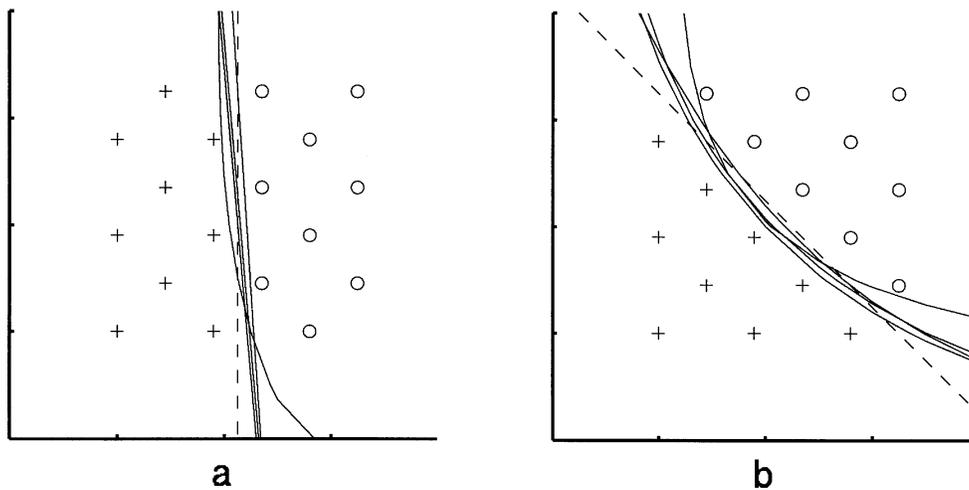


Figure 3. Stimuli, experimenter-defined, and “most parsimonious” category boundaries for each observer in the (a) selective attention experiment and (b) linear integration experiment. The plus signs denote “A” stimuli, and the circles denote “B” stimuli. The dashed line denotes the experimenter-defined bound, and the solid lines denote the “most parsimonious” bound.

be slower than correct responses. The results from the selective attention experiment neither support nor provide strong evidence against this prediction. It is worth mentioning that the predictions of the RT–distance hypothesis and the finding of errors being slower than correct responses are somewhat counterintuitive since we would expect a decision process that continues longer to be more accurate than a decision process that does not continue for as long. The RT–distance hypothesis predicts this counterintuitive result because percepts leading to errors will, on average, lie closer to the category boundary than will percepts that lead to correct responses.

Why might the correct-response and error mean RTs support the RT–distance hypothesis in the linear integration experiment and not in the selective attention experiment? There are at least two possible explanations. First, there is evidence that task difficulty affects the prevalence of “fast guessing,” with easy tasks yielding more fast guesses than difficult tasks. A robust result in the two-choice RT literature is that errors are often faster than correct responses when the discrimination is easy, but they are slower when the discrimination is difficult (see Link & Heath, 1975, Luce, 1986, Ratcliff, 1978, Townsend & Ashby, 1983, and Vickers, 1979, for reviews of this literature). One possible explanation of this result is that, when the discrimination is very easy, any processing of the stimulus leads to a correct response. As a result, when observers are pressed to respond even more quickly, they can do so only by ignoring all stimulus information on some proportion of trials and by guessing. This “fast-guess” model of the speed–accuracy tradeoff predicts that error RTs will be faster than correct-response RTs (Ollman, 1967; Yellott, 1968). Notice that the fast errors do not result because errors are processed faster than correct responses, but rather because many errors result from guesses that happen to be fast but incorrect.

When the discrimination is relatively difficult, an emphasis on speed causes observers to respond on the basis of partial processing. No (or very little) fast guessing occurs, but stimuli are not processed completely. Since all stimuli are processed, percepts leading to errors will, on average, lie closer to the category boundary than will percepts that lead to correct responses, with the result that error RTs are longer than correct-response RTs. In this case, the RT–distance hypothesis is supported. These findings are relevant, and they might explain the difference in correct-response and error mean RT orderings across experiments, because the selective attention experiment should have been an easier task to perform than the linear integration experiment. In the selective attention experiment, only one stimulus component was relevant, and all processing could be focused on that component, whereas, in the linear integration experiment, both stimulus components had to be processed. In addition, the fact that the stimuli were composed of separable dimensions should have made this focusing operation especially easy, and the integration process more difficult. This fast-guess hypothesis is speculative at this point, but it provides an interesting account of the data. Clearly, more work is necessary to rigorously test this hypothesis (see Townsend & Ashby, 1983, pp. 263–271, for a review of more rigorous methods of testing the fast-guess hypothesis).

A second possibility is that the response bias observed in the accuracy data might account for some of the cases in which error mean RTs were shorter than correct-response mean RTs. The logic is as follows. Suppose an observer is biased toward response “A.” In this case, the observer will be fast and accurate when response “A” is correct and will be slow and less accurate when response “B” is correct. In addition, this observer will be slow when incorrectly responding “B” to an “A” stimulus but should be fast when

incorrectly responding “A” to a “B” stimulus. In short, such a bias predicts that errors should be relatively slow and correct responses fast for the “biased” category, whereas errors should be fast and correct responses slow for the “unbiased” category. To test this hypothesis, we examined the correct-response and error mean RTs for the biased and unbiased categories for each case in which a bias was observed in the accuracy data. Because there were nine Category A stimuli and nine Category B stimuli, there were nine tests of this hypothesis for the biased and unbiased categories. Recall that there were six cases in which a response bias was observed. Observer 1 showed a bias toward response “A” in the linear integration experiment, Observer 2 showed a bias toward response “B” in both experiments, Observer 3 showed a bias toward response “A” in both experiments, and Observer 4 showed a bias toward response “B” in the linear integration experiment. The data in Tables 3 and 4 provide strong support for this “response bias” hypothesis. For five of the six cases in which a bias existed, six of nine stimuli showed the predicted ordering for the biased category (i.e., longer error mean RT than correct-response mean RT). In the final case, five of seven stimuli showed the predicted ordering (for two stimuli, there were no errors). In five of the six cases, nine of nine stimuli showed the predicted ordering for the unbiased category (i.e., shorter error mean RT than correct-response mean RT). In the final case, eight of nine stimuli showed the predicted pattern. It is important to note that, although the RT–distance hypothesis predicts that responses (both correct and error) to the favored category will be faster than responses to the nonfavored category, it still always predicts that, for any particular stimulus, errors should be slower than corrects. Thus, the response bias hypothesis tested here is fundamentally incompatible with the RT–distance hypothesis.

Ratcliff and colleagues (Ratcliff & Rouder, 1997; Ratcliff, Van Zandt, & McKoon, 1997) recently developed and tested a diffusion model that predicts the pattern of results observed in these experiments. The critical factor in this model is that there is variability across trials both in the starting point of the diffusion process and in the mean drift rate. Variability in the starting point leads to fast errors, and variability in the mean drift rate leads to slow errors. When applied to conditions that vary in difficulty, one or the other factor tends to dominate the predicted RT. In line with the present results, when the discrimination is easy, errors are predicted to be faster than correct responses, and when the discrimination is difficult, errors are predicted to be slower than correct responses.

**Correct-response mean RT.** A robust finding in previous research is that correct-response mean RT decreases as distance to the category boundary increases (Bornstein & Monroe, 1980; Cartwright, 1941). In line with previous research, this RT–distance hypothesis finds support across experiments, observers, and categories (see Table 5). In every case, mean RT decreased as distance to bound increased.<sup>5</sup>

To test the predicted correct-response mean RT orderings more rigorously, *t* tests were conducted between every

possible pair of stimuli (see Equation 1). Comparisons can be classified into four types: one-step near, one-step far, two-step, and parallel. The one- and two-step comparisons have already been defined. The parallel comparisons are those between stimuli that were the same distance from the experimenter-defined decision bound. Statistically significant one- and two-step mean RT differences would provide evidence in support of the RT–distance hypothesis, whereas statistically nonsignificant parallel RT differences provide support for the hypothesis. To be conservative in our statistical analyses, we used different significance levels for the one- and two-step comparisons and the parallel comparisons. Specifically, we set  $\alpha = .01$  for the one- and two-step comparisons, and  $\alpha = .30$  for the parallel comparisons.

These data from these analyses are in Table 6, which presents the percentages of comparisons that provide support for the RT–distance hypothesis. Also included are the mean RT differences associated with each series of statistical tests, as well as the number of *t* tests conducted for each type of comparison. The results can be summarized as follows. First, restricting attention to the one- and two-step comparisons, a large percentage of the tests supported the RT–distance hypothesis. On the basis of the averaged data from both experiments, 100% of the two-step comparisons and well over 70% of the one-step comparisons supported the RT–distance hypothesis. In addition, notice that the mean RT differences are relatively large, ranging from 14 to 64 msec. The data from the parallel comparisons provided somewhat less support. Even so, nearly 60% and just over 70% of the parallel tests supported the RT–distance hypothesis in the selective attention and linear integration experiments, respectively. Note also that the parallel mean RT differences are small (9 and 21 msec). The latter result was not completely unexpected given the fact that the best-fitting decision bounds deviated slightly from the experimenter-defined bound. The “parallel” stimuli are equidistant only from the experimenter-defined bound. If the observer’s bound differs from the experimenter-defined bound, then many of these stimulus pairs will not be equidistant. An examination of Figure 3 suggests that there were small deviations from the experimenter-defined boundary. Even so, a large percentage of the parallel tests supported the distance to bound hypothesis.

**Cumulative RT distributions.** A stronger test of the RT–distance hypothesis requires a comparison of the cumulative RT distributions (see Equation 2). Specifically, the RT–distance hypothesis is supported if  $F_j(t) \geq F_i(t)$  for all  $t > 0$  and Stimuli *i* and *j* for which  $d_{iB} < d_{jB}$ , where  $d_{iB}$  is the distance from Stimulus *i* to the category boundary. To test this hypothesis, Kolmogorov–Smirnov tests were conducted using the same strategy described for the mean RT comparisons. Specifically, all possible one-step, two-step, and parallel comparisons were tested using the significance levels outlined above.

These analyses are presented in Table 7 using the same format as that used in Table 6. The results mirror those for mean RT. Strong support was found for the RT–distance

**Table 6**  
**Percentages of Correct-Response (CR) Mean RT Comparisons That Differed Significantly**  
**(Based on *t* Tests) and the Average (Av) Mean RT Differences for These Comparisons (in Milliseconds)**

Comparison	No. of <i>t</i> Tests	Observer								Average	
		1		2		3		4		CR	Av
		CR	Av	CR	Av	CR	Av	CR	Av		
Selective Attention Experiment											
One-step far	18	83	12	67	21	89	13	72	10	78	14
One-step near	18	50	15	83	35	89	20	89	17	78	22
Two-step	18	100	23	100	48	100	30	100	24	100	31
Parallel	18	44	6	78	134	56	6	56	7	59	9
Linear Integration Experiment											
One-step far	6	67	11	67	21	83	16	67	14	71	16
One-step near	30	100	38	100	69	100	48	100	47	100	51
Two-step	10	100	46	100	87	100	62	100	59	100	64
Parallel	26	81	14	73	29	69	25	65	17	72	21

Note—For the one- and two-step comparisons, the significance level was set to  $\alpha = .01$ . For the parallel comparisons, the significance level was set to  $\alpha = .30$ .

hypothesis in both experiments for all observers and for the one-step, two-step, and parallel comparisons. For illustrative purposes, the cumulative RT distributions for one small-, medium-, and large-distance stimulus are presented in Figure 4 for a representative observer (Observer 2). For the selective attention experiment (Figure 4a), the presented stimuli are 7, 10, and 8. For the linear integration experiment (Figure 4b), the analogous stimuli are 1, 4, and 8. Notice that the cumulative RT distributions are clearly ordered in the way predicted by the RT-distance hypothesis.

**RT hazard function ordering.** The hazard functions were estimated using the random smoothing technique of Miller and Singpurwalla (1977; see also Ashby et al., 1993). Unlike the analyses of the mean RT and cumulative RT distributions, no well-established statistical test exists for determining whether two hazard functions are ordered. Thus, we simply plotted all pairs of hazard functions for the one- and two-step comparisons and inspected them visually. Although we have no way of quantifying these analyses, in general, the hazard function orderings supported the RT-distance hypothesis. For example, Figure 5 depicts the hazard functions for the same stimuli presented in Figure 4. Although there are some violations of the predicted ordering, in general, the hazard functions clearly appear to be ordered by distance to the bound.

**RT hazard function shape.** The shape of the hazard functions is also informative. Notice (in Figure 5) that the hazard functions for the stimuli farthest from the bound are more “peaked” than are the functions for the stimuli close to the bound. Although differing in magnitude, this pattern held up for all stimuli and observers. Interestingly, this same pattern has been observed in simple detection (Burbeck & Luce, 1982), subitizing (Balakrishnan & Ashby, 1992), and memory scanning (Ashby et al., 1993). Specifically, in each case, the hazard function is peaked for the “easier” trials and is less peaked for the “difficult” trials. This is especially interesting because of the large RT variation across these tasks. For example, mean RT on a difficult

detection trial (i.e., low-intensity stimulus) is typically less than mean RT on an easy memory scanning trial (i.e., small memory set). Yet, within a task, the same change in the hazard function occurs with changes in the difficulty of the trials. This remarkable similarity across such different tasks suggests a possible common mechanism that may operate in virtually all perceptual decision-making tasks.

The nonmonotonicity of the hazard functions associated with the stimuli that are farthest from the bound rules out a large class of serial processing models (see, e.g., Ashby et al., 1993). The RT hazard function at time *t* gives the likelihood that a response will be made in the next instant, given that one has not already been made. In a serial process, as time increases, the number of stages remaining uncompleted tends to decrease. As a result, most serial models predict that the hazard function steadily increases with time. Such models cannot account for the hazard functions shown in Figure 5.

**RT likelihood ratios.** A test of the monotonicity of likelihood ratios is the strongest test of stochastic domi-

**Table 7**  
**Percentages of Correct-Response Cumulative**  
**RT Comparisons That Differed Significantly**  
**(Based on Kolmogorov–Smirnov [K–S] Tests)**

Comparison	No. of K–S Tests	Observer				Average
		1	2	3	4	
Selective Attention Experiment						
One-step far	18	61	94	78	39	68
One-step near	18	44	83	67	83	69
Two-step	18	100	100	100	100	100
Parallel	18	33	61	39	22	38
Linear Integration Experiment						
One-step far	6	67	100	67	67	75
One-step near	30	100	100	100	100	100
Two-step	10	100	100	100	100	100
Parallel	26	65	69	88	50	68

Note—For the one- and two-step comparisons, the significance level was set to  $\alpha = .01$ . For the parallel comparisons, the significance level was set to  $\alpha = .30$ .

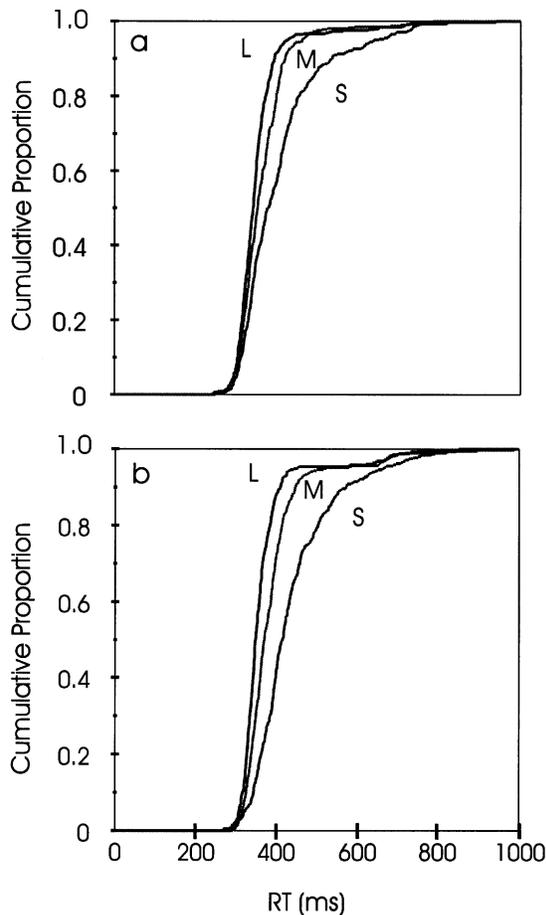


Figure 4. Representative short-, medium-, and long-distance-to-bound cumulative RT distributions from Observer 2 for the (a) selective attention experiment and (b) linear integration experiment. In the selective attention experiment, the plots are for Stimuli 7, 8, and 10. In the linear integration experiment, the plots are for Stimuli 1, 4, and 8.

nance of stochastic dominance considered in this article. From a statistical standpoint, likelihood ratios are difficult to estimate. Fortunately, however, there is a statistically reliable method of determining whether likelihood ratios are nondecreasing that does not require an estimate of the likelihood ratio (Ashby et al., 1993). A well-known result in signal detection theory states that the likelihood ratio (the ratio of the signal-plus-noise over the noise distributions) is an increasing function of the sensory variable, if and only if the ROC curve is concave<sup>6</sup> (Laming, 1973; Peterson, Birdsall, & Fox, 1954). The survivor function is defined as one minus the cumulative distribution function,  $1 - F(t)$ , so an ROC curve is a plot of the survivor function of the signal-plus-noise distribution [i.e.,  $P(\text{hit})$ ] against the survivor function of the noise distribution [i.e.,  $P(\text{false alarm})$ ]. Therefore, an alternative method for determining whether the likelihood ratio,  $l(t) = f_i(t)/f_j(t)$ , is increasing in  $t$  is to plot the survivor function of Stimulus  $i$

against the survivor function of Stimulus  $j$  and check whether the resulting function is concave (see Ashby et al., 1993, for more details). Unfortunately, we know of no statistical test for determining whether a function is concave, so visual inspection was employed. Overall, there was no evidence that concavity was violated. This is seen clearly in Figure 6, which shows the RT-ROC curves for the same stimuli used in Figures 4 and 5.

### Summary

The stochastic dominance tests indicate that the distance from the stimulus to the category boundary strongly affects the time it takes for the observer to respond. Specifically, the tests demonstrate that stimuli that are close to the category boundary yield RTs that are stochastically greater than RTs for stimuli that are farther from the category boundary. These findings provide strong support for the RT-distance hypothesis (Ashby & Maddox, 1994; Maddox & Ashby, 1996). The RT-distance hypothesis was supported in two qualitatively different types of categorization problems: one in which the experimenter-defined

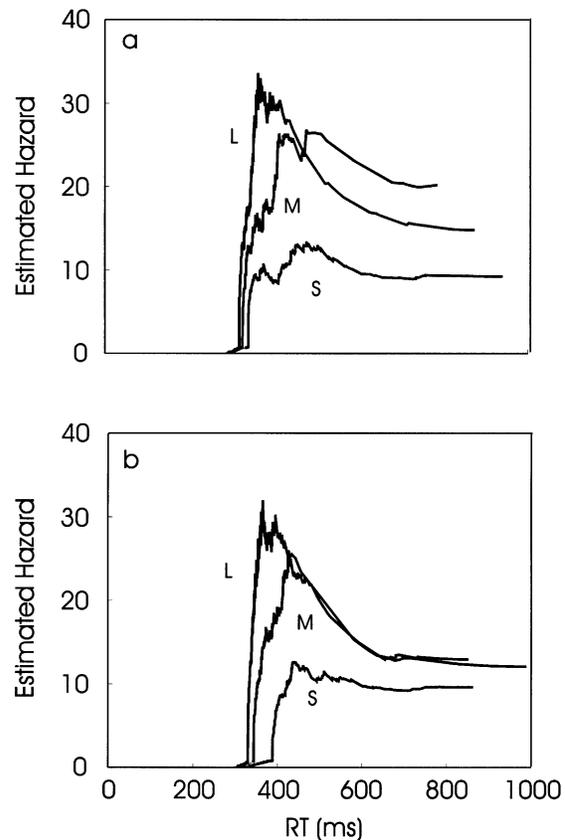


Figure 5. Representative short-, medium-, and long-distance-to-bound hazard functions from Observer 2 for the (a) selective attention experiment and (b) linear integration experiment. In the selective attention experiment, the plots are for Stimuli 7, 8, and 10. In the linear integration experiment, the plots are for Stimuli 1, 4, and 8.

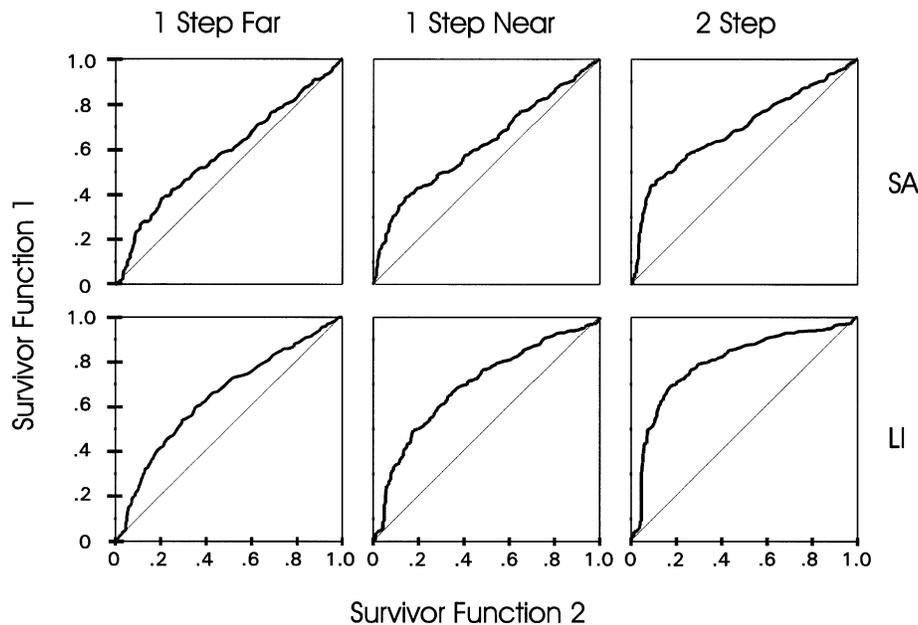


Figure 6. Representative ratio of the cumulative distribution plots for a one-step far, one-step near, and two-step comparison from Observer 2 for the selective attention and linear integration experiments. In the selective attention experiment, the plots are for Stimuli 8 and 10 (one-step near), 7 and 10 (one-step far), and 7 and 8 (two-step). In the linear integration experiment, the plots are for Stimuli 4 and 8 (one-step near), 1 and 4 (one-step far), and 1 and 8 (two-step).

strategy was to ignore a stimulus component (i.e., attend selectively), and another in which both components needed (approximately) equal attention. In both cases, however, the experimenter-defined decision bound was approximately linear. In contrast to predictions of the RT–distance hypothesis, error mean RTs were not always longer than correct mean RTs, especially in the selective attention experiment. One possible explanation is that the relatively easy selective attention experiment led to a greater number of fast guesses. A second possibility is that a response bias existed for many observers that led to faster responding for one category label over the other. This hypothesis provided a good account of the data, and it appears to account for many of the violations of the RT–distance hypothesis.

To extend our analyses of RT distributions even further, we had observers participate in a categorization problem with a highly nonlinear category boundary. We turn to this experiment now.

### Nonlinear Integration Experiment

Because the category boundary is nonlinear, there is no simple method for classifying the stimuli into small-, medium-, and large-distance categories, as was done in the selective attention and linear integration experiments. However, several pairs of stimuli clearly differed in distance to bound. Thus, we will focus primarily on these pairs of stimuli. For example, in Category A, Stimulus 1 and Stimulus 18 were farther from the bound than were Stimuli 4, 7, 12, 13, 15, 16, and 17. Thus, the stochastic dominance tests will be performed between Stimulus 1 and each of these seven

other stimuli and between Stimulus 18 and each of these other stimuli. In Category B, Stimulus 3, Stimulus 5, and Stimulus 8 were farther from the bound than were Stimuli 2, 6, 9, 10, 11, and 14. Thus, the stochastic dominance tests will be performed between each one of these relevant pairs. In total, we identified 32 pairs of stimuli to test.

### Accuracy Analysis and Observer's Category Boundary

Table 8 presents the accuracy rates for each stimulus by observer for the nonlinear integration experiment. Figure 7 presents the experimenter-defined and best-fitting quadratic category boundary. As in the previous experiments, the accuracy rates increased with distance to bound. In addition, the best-fitting boundary was similar to the experimenter-defined boundary and preserved the distance-to-bound relations. Interestingly, there were a few stimuli that appeared to give the observers some trouble. For example, Observer 3 frequently classified Stimuli 4, 10, and 12 into the incorrect category. The accuracy rates for these stimuli for Observer 3 were 33%, 48%, and 34%, respectively. The best-fitting quadratic boundary for Observer 3 predicted accurately the poor performance for Stimuli 4, 10, and 12, yielding predicted accuracy rates of 34%, 47%, and 37%, respectively.

### Response Time Analyses

**Correct-response and error mean RT.** Table 8 displays the correct-response and error mean RTs for each observer. As in the previous experiments, correct-response

**Table 8**  
**Accuracy Rates (Acc) and Correct-Response (CR) and Error (E) Mean RTs**  
**by Stimulus and Observer for the Nonlinear Integration Experiment**

Stimulus	Observer									Average		
	1			2			3			Acc	CR	E
	Acc	CR	E	Acc	CR	E	Acc	CR	E			
1	97.49	481	472	95.04	437	441	93.20	465	465	95.16	459	457
2	89.86	503	610	95.59	400	478	97.21	429	482	94.26	443	556
3	99.79	436	387	99.56	380	715	99.59	407	612	99.65	408	608
4	61.81	600	593	48.58	519	444	33.04	513	471	47.77	552	492
5	98.83	477	566	98.23	379	403	98.87	402	486	98.64	418	471
6	99.06	459	499	97.39	401	440	96.15	411	501	97.54	424	479
7	98.32	505	470	96.23	446	392	94.39	460	442	96.29	470	429
8	93.16	563	612	98.25	398	538	94.93	436	471	95.45	464	550
9	95.50	525	661	97.34	403	653	98.26	406	563	97.02	446	639
10	68.88	625	584	59.30	502	492	47.91	517	481	59.09	556	512
11	83.48	577	636	94.85	435	533	93.36	445	490	90.54	482	583
12	64.22	647	581	72.12	540	429	34.22	526	455	57.15	577	484
13	97.49	484	652	97.39	431	512	97.78	446	552	97.56	453	571
14	78.43	585	508	80.24	502	491	61.44	515	503	73.59	537	502
15	81.22	592	637	87.45	498	469	58.88	502	495	76.85	532	528
16	98.38	473	533	92.44	475	542	92.94	458	604	94.64	469	569
17	98.72	482	452	94.56	465	482	89.62	472	470	94.37	473	472
18	98.63	472	481	99.77	431	298	94.03	462	438	97.34	455	441
Average	88.85	520	586	88.94	441	469	82.38	449	479	86.72	471	506

Note—Correct-response and error mean RTs are in milliseconds.

mean RT was longest near the decision bound and fell off monotonically with distance to bound. To test the RT–distance hypothesis, we performed *t* tests on the correct-response and error mean RTs for each stimulus and observer. Averaged across observers and stimuli, 37% of the correct-response mean RTs were significantly faster than the error mean RTs (in support of the RT–distance hypothesis), and 17% of the correct-response mean RTs were significantly slower than the error mean RTs (providing evidence against the RT–distance hypothesis). Forty-six percent of the tests were not significant and, thus, did not provide support for or evidence against the RT–distance hypothesis. Averaged across stimuli, the error mean RTs were longer than the correct-response mean RTs for each observer. The difference was quite large (average error minus correct-response mean RT = 35 msec) and was somewhat larger than in the linear integration experiment (average error minus correct-response mean RT = 25 msec) and much larger than in the selective attention experiment (average error minus correct-response mean RT = –3 msec). A similar pattern holds when we examine the percentages of *t* tests that provide statistically significant support for the RT–distance hypothesis: 37%, 15%, and 7% in the nonlinear integration, linear integration, and selective attention experiments, respectively. Notice that these data provide support for the hypothesis suggested earlier that task difficulty increases the difference between correct-response and error mean RT. Although the correct-response and error mean RT data, averaged across stimuli, provide strong support for the RT–distance hypothesis, there were still several stimuli for which the ordering was violated. When discussing the selective attention and linear integration data, we hypothesized that a response bias might account for some of these violations. Further ex-

amination of the nonlinear integration data suggests that a response bias might also be operative in this experiment. Although Observer 1 showed no bias, Observers 2 and 3 showed a strong bias to respond “B.” Thus, when response “B” is the correct response, we expect correct responses to be fast and errors to be slow. When response “A” is the correct response, we expect correct responses to be slow and errors to be fast. In support of this hypothesis, for seven of the nine Category B stimuli, the correct-response mean RT was shorter than the error mean RT. Analogously, for five of nine (Observer 2) and seven of nine (Observer 3) Category A stimuli, the correct-response mean RT was longer than the error mean RT. Despite the possible response bias, taken as a whole, the correct-response and error mean RTs supported the RT–distance hypothesis.

To determine whether the 32 pairs of stimuli that differ in distance to bound yielded mean RT differences in the direction predicted by the RT–distance hypothesis, *t* tests were performed. For Observers 1, 2, and 3, respectively, 81%, 81%, and 56% of the tests provided support for the RT–distance hypothesis.

**Cumulative RT distributions.** The same stimulus pairs were used to test the stronger form of stochastic dominance based on the cumulative RT distributions. For Observers 1, 2, and 3, respectively, 88%, 84%, and 66% of the Kolmogorov–Smirnov tests provided support for the RT–distance hypothesis. Note that these percentages were higher than for the analogous tests on the mean RTs. Since an ordering of the cumulative distribution functions implies an ordering of the mean RTs, these results, therefore, reinforce our conclusion that the mean RT orderings are reliable. A plot of the cumulative RT distributions for Stimuli 1 and 4 for a representative observer (Observer 2) are depicted in Figure 8a.

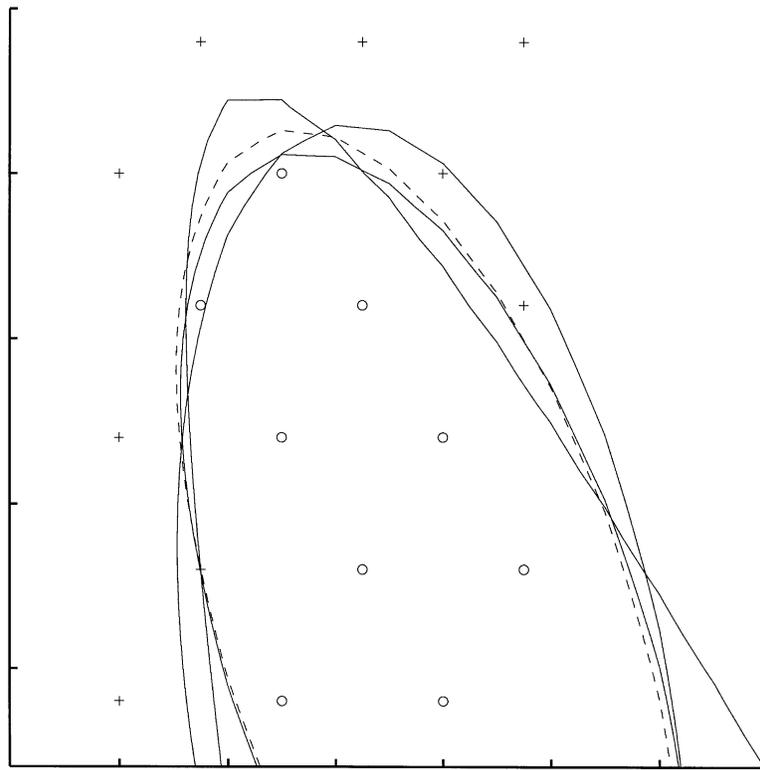


Figure 7. Stimuli, experimenter-defined, and best-fitting quadratic category boundaries for each observer in the nonlinear integration experiment. The plus signs denote "A" stimuli, and the circles denote "B" stimuli. The dashed line denotes the experimenter-defined bound, and the solid lines denote the "most parsimonious" bound.

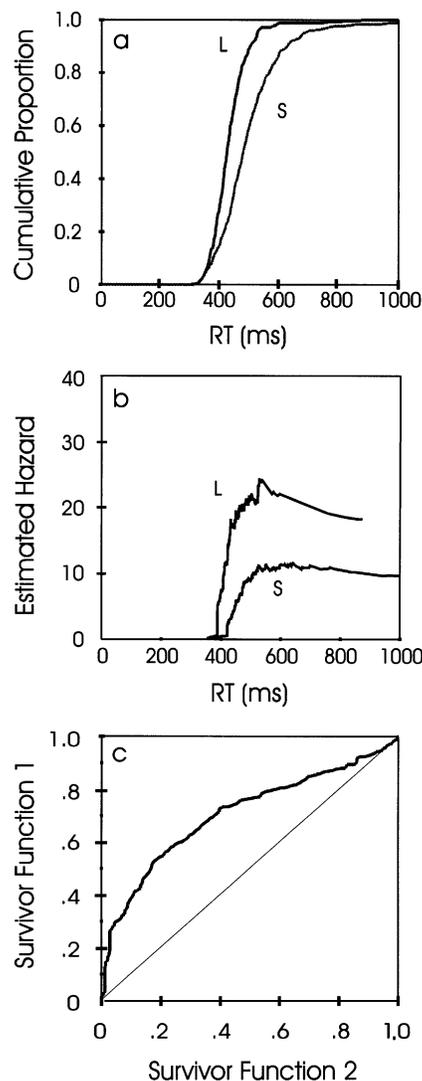
**RT hazard functions and RT likelihood ratios.** The RT hazard functions and likelihood ratios in the nonlinear integration experiment mimicked those in the two linear experiments. In general, on the basis of visual inspection, the hazard functions were ordered in the direction predicted by the RT–distance hypothesis. In addition, the hazard functions were again peaked for stimuli that were farthest from the decision bound. Finally, although some violations of concavity existed, the RT–ROC curves were generally concave, which supports the hypothesis that the likelihood ratios increased with RT. Figures 8b and 8c depict the hazard functions and the RT–ROC for Stimuli 1 and 4 from Observer 2.

#### GENERAL DISCUSSION AND CONCLUSIONS

The goal of this study was to collect a rich set of categorization RT data that could be used to test developing models of categorization RT. The observers participated in three categorization experiments that differed qualitatively in the nature of the experimenter-defined category boundary. In the selective attention experiment, the observer was required to attend selectively to one dimension of the stimulus while ignoring the other dimension. In the linear

integration experiment, the observer was required to attend (approximately) equally to both stimulus dimensions and integrate the stimulus information in a linear fashion. In the nonlinear integration experiment, the observer was required to attend to both stimulus dimensions and integrate the stimulus information in a nonlinear fashion. The observers completed several sessions in each experiment, resulting in a large number of repetitions for each individual stimulus. These large sample sizes made it possible to obtain accurate estimates of the RT distributions, thereby allowing us to examine the effects of many important factors on categorization RT.

A robust finding in the empirical literature is that correct-response mean RT decreases with the distance between the exemplar and the category boundary (e.g., Bornstein & Monroe, 1980; Cartwright, 1941; see also Ashby et al., 1994). This RT–distance hypothesis found support in the data from all observers across all three experiments both at the relatively weak level of correct-response mean RT and at higher distributional levels (i.e., at the level of the cumulative RT distribution, the RT hazard function, and the RT likelihood ratio). In addition, we found strong evidence of RT invariance for all stimuli the same distance from the boundary (and within the same category). Thus, we found no evidence that position within the category



**Figure 8.** Representative (a) cumulative RT distributions, (b) hazard functions, and (c) ratio of the cumulatives for Observer 2 from the nonlinear integration experiment. The plots are for Stimuli 1 and 4.

structure had any effect on RT (once distance to bound was controlled). Among the important implications of this result is that our data showed no signs of a similarity or typicality effect. For example, Figures 1 and 2 indicate that in the linear integration experiment, Stimulus 8 was more similar and more typical to the Category A exemplars than were Stimuli 3 or 13 (by any currently popular similarity and typicality measures), although distance to the boundary was the same for these three stimuli. Thus, if there was a tendency for the observers to respond "A" more quickly to stimuli that were more typical of Category A or more similar to the Category A exemplars, then RT to Stimulus 8 would have been less than RT to Stimuli 3 or 13. We found no evidence of such differences.

In its most popular forms, the RT–distance hypothesis predicts error RTs to be larger than correct-response RTs (Thomas, 1971; see also Thomas & Myers, 1972). This prediction was supported in the linear integration and nonlinear integration experiments; however, in the selective attention experiment, correct and error RTs were approximately equal. The selective attention task is the "easiest" task, in the sense that the observer need only attend to one of two separable stimulus components. The nonlinear integration task is the most difficult because both stimulus components must be attended, and information about these components must be integrated in a nonlinear fashion. The linear integration task is intermediate in difficulty because it requires information integration, but integration in a linear fashion. Two reasonable, but tentative, explanations are offered for these results. One possibility is that the relatively easy selective attention experiment led to a greater number of fast guesses than did the more difficult integration experiments and thus led to faster error RTs. A second possibility is that a response bias existed for many observers that led to faster responding for one category label over the other. Both hypotheses are reasonable, and they appear to account for many of the violations of the RT–distance hypothesis.

Another interesting aspect of the present data was the effect of distance to the boundary on the shape of the RT hazard functions. Specifically, the hazard functions for the stimuli farthest from the bound were more "peaked" than were the functions for the stimuli close to the bound. Interestingly, this same pattern has been observed in simple detection (Burbeck & Luce, 1982), subitizing (Balakrishnan & Ashby, 1992), and memory scanning (Ashby et al., 1993). In each case, the hazard function is peaked for the easier trials and is less peaked for the difficult trials. This similarity across such different tasks suggests a possible common mechanism that may operate in virtually all perceptual decision-making tasks. In addition, the nonmonotonicity of the hazard functions associated with the stimuli that are farthest from the bound raised a large class of serial processing models (see, e.g., Ashby et al., 1993).

To our knowledge, this is the first categorization study with sample sizes large enough for accurate estimation of the RT distributions from individual observers. The rich data set that resulted can be used for quantitative testing of alternative categorization models, but an analysis of those data also uncovered a number of qualitative results that any serious model of categorization RT must predict. In particular, any viable categorization model must make several predictions: (1) RT is faster for stimuli farther from the category boundary, and this stochastic dominance holds all the way up to the level of the RT likelihood ratio. (2) RT is invariant for all stimuli the same distance from the category boundary, at least in experiments where the stimuli are all presented with equal frequency. In particular, this implies that, in such experiments, similarity and typicality have no fundamental effect on RT. (3) As in two-choice discrimination, the relation between correct and

error RT depends on task difficulty. When the difficulty is high, errors are slower than correct responses, whereas this difference disappears when difficulty is low. (4) Small, consistent response biases appear to have a large effect on the relation between correct and error RT. (5) Categorization RT hazard functions are qualitatively similar to hazard functions observed in detection, subitizing, and memory scanning experiments. Specifically, the hazard functions are ordered by trial difficulty (i.e., by distance to boundary), they have flat tails, they are peaked on easy trials (i.e., for stimuli far from the boundary), and they are increasing on difficult trials (i.e., for stimuli near the boundary).

These results present an immediate challenge to existing categorization RT models, and they should serve as a valuable guide to researchers developing new theories.

#### REFERENCES

- ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.
- ASHBY, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G., BOYNTON, G., & LEE, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11-27.
- ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150-172.
- ASHBY, F. G., & LEE, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369-399). Amsterdam: Elsevier, North-Holland.
- ASHBY, F. G., & MADDOX, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 598-612.
- ASHBY, F. G., & MADDOX, W. T. (1991). A response time theory of perceptual independence. In J. P. Doignon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 389-414). New York: Springer-Verlag.
- ASHBY, F. G., & MADDOX, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 50-71.
- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between exemplar, prototype, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- ASHBY, F. G., & MADDOX, W. T. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423-466.
- ASHBY, F. G., & MADDOX, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Handbook of perception and cognition: Vol. 3. Judgement, decision making, and measurement* (pp. 251-301). New York: Academic Press.
- ASHBY, F. G., TEIN, J., & BALAKRISHNAN, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, **37**, 526-555.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- BALAKRISHNAN, J. D., & ASHBY, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, **54**, 80-90.
- BORNSTEIN, M. H., & MONROE, M. D. (1980). Chromatic information processing: Rate depends on stimulus location in the category and psychological complexity. *Psychological Research*, **42**, 213-225.
- BURBECK, S. L., & LUCE, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, **32**, 117-133.
- CARTWRIGHT, D. (1941). Relation of decision time to the categories of response. *American Journal of Psychology*, **54**, 174-196.
- ESTES, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
- GARNER, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- GARNER, W. R., & FELFOLDY, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, **1**, 225-241.
- GEISLER, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, **96**, 267-314.
- GREEN, D. M., & SWETS, J. A. (1967). *Signal detection theory and psychophysics*. New York: Wiley.
- HOMA, D., DUNBAR, S., & NOHRE, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 444-458.
- HUTTENLOCHER, J., HEDGES, L. V., & DUNCAN, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, **98**, 352-376.
- LAMING, D. (1968). *Information theory of choice reaction time*. London: Academic Press.
- LAMING, D. (1973). *Mathematical psychology*. New York: Academic Press.
- LINK, S. W., & HEATH, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77-105.
- LUCE, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- MADDOX, W. T. (1995). Baserate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1-14.
- MADDOX, W. T. (in press). On the dangers of averaging across observers when comparing decision bound and generalized models of categorization. *Perception & Psychophysics*.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.
- MADDOX, W. T., & ASHBY, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 301-321.
- MILLER, D. R., & SINGPURWALLA, N. D. (1977). *Failure rate estimation using random smoothing* (Tech Rep. AD-A040999/5ST). National Technical Information Service.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- NOSOFSKY, R. M. (1992). Exemplar-based approach to relating categorization, identification and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-394). Hillsdale, NJ: Erlbaum.
- NOSOFSKY, R. M., CLARK, S. E., & SHIN, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 282-304.
- NOSOFSKY, R. M., & PALMERI, T. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.
- OLLMAN, R. (1967). Fast guesses in choice reaction time. *Psychonomic Science*, **6**, 192-208.
- PETERSON, W. W., BIRDSALL, T. G., & FOX, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, **4**, 171-212.
- RATCLIFFE, R. (1978). A theory of memory retrieval. *Psychological Review*, **71**, 59-108.
- RATCLIFFE, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.

- RATCLIFF, R., & ROUDER, J. N. (1997). *Modeling response times for two-choice decisions*. Manuscript submitted for publication.
- RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1997). *Connectionist and diffusion models of reaction time*. Manuscript submitted for publication.
- REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382-278.
- SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54-87.
- THOMAS, E. A. C. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, *8*, 303-332.
- THOMAS, E. A. C., & MYERS, J. L. (1972). Implications of listing data for threshold and non-threshold models of signal detection. *Journal of Mathematical Psychology*, *9*, 253-285.
- TOWNSEND, J. T. (1991). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, *106*, 551-567.
- TOWNSEND, J. T., & ASHBY, F. G. (1978). Methods of modeling capacity in simple processing systems. In N. J. Castellan, Jr., & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 199-239). Hillsdale, NJ: Erlbaum.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- VICKERS, D. (1979). *Decision processes in visual perception*. New York: Academic Press.
- WELFORD, A. T. (1968). *Fundamentals of skill*. London: Academic Press.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W. H. Freeman.
- YELLOTT, J. I., JR. (1968). Correction for guessing in choice reaction time. *Psychonomic Science*, *8*, 321-322.

#### NOTES

1. Ratcliff (1979) outlined a method for using group data to estimate higher level statistics. This approach is useful when data from many observers is available, but there are only a few stimulus presentations per observer. Ratcliff's (1979) approach would be especially useful when learning is of direct interest.

2. RT variability could be caused by one of a number of factors. With simple stimuli, such as those used in the present study, it is likely that there is trial-by-trial variability in the perceptual effect for repeated presentations of the same stimulus (e.g., Ashby & Lee, 1993; Geisler, 1989; Green & Swets, 1967). In addition, variation in the time to initiate the appropriate motor program is likely. For the present purposes, it is only important to acknowledge that RT variability exists.

3. The number of sessions to be considered as practice was determined for each observer by examining the overall accuracy rate and correct-response mean RT across sessions. Across the first few sessions, the accuracy rate tended to increase and the correct-response mean RT tended to decrease. This is most likely due to learning of the exemplars and their category mappings. Across the remaining sessions, the accuracy rate and the correct-response mean RT was fairly constant. Because the focus of this research was on asymptotic categorization performance, the early learning sessions were considered practice and were excluded from subsequent analyses.

4. Thomas (1971) showed that if the RT-distance hypothesis holds, and the hazard function of the perceptual distribution is increasing, then median incorrect RT will be greater than median correct RT. The hazard functions of many well-known probability distributions are increasing (e.g., gamma, logistic, ex-Gaussian, Rayleigh). This includes the normal distribution, which has been the most common distributional assumption in signal detection theory (e.g., Ashby & Townsend, 1986; Green & Swets, 1967).

5. Although the distances between the small- and medium-distance stimuli and between the medium- and large-distance stimuli were identical, a close examination of Tables 3-5 suggests that the differences in performance for small- and medium-distance stimuli and for medium- and large-distance stimuli were not equal. Performance for the small- and medium-distance stimuli differed substantially, whereas performance for the medium- and large-distance stimuli was very similar. This finding was not unexpected. In fact, only a linear function relating distance to RT would predict equal performance differences for small- versus medium-distance stimuli and for medium- versus large-distance stimuli.

A detailed discussion of this issue is beyond the scope of this article; however, the interested reader is directed to Ashby and Maddox (1991, 1994) for details.

6. A curve is concave if any two points on the curve can be connected by a line segment that lies completely below the curve.

#### APPENDIX

In this appendix, we outline the procedure used to determine each observer's best-fitting linear and quadratic category boundary. Two decision bound models of categorization were applied to the data from each observer in the selective attention, linear integration, and nonlinear integration experiments. These were the general linear classifier and the general quadratic classifier. In short, the general linear classifier assumes the observer uses a linear decision boundary but not necessarily the experimenter-defined decision boundary, and the general quadratic classifier assumes the observer uses a quadratic decision boundary. The details of these models are described fully in many other articles (e.g., Ashby, 1992a, 1992b; Ashby & Maddox, 1993; Maddox, 1995; Maddox & Ashby, 1993). In applying these models to the data, some assumptions must be made regarding the relation between the perceptual space and the physical space. Following Maddox and Ashby (1993), we assumed that the mean perceptual effect for Stimulus  $i$  was equal to the coordinates of Stimulus  $i$  in the physical space. In addition, the perceptual covariance matrix was assumed to be a scalar multiple of the identity matrix and was assumed to be bivariate normal. Using these same assumptions, Maddox and Ashby (1993) successfully accounted for categorization accuracy in a series of experiments that utilized these same stimulus dimensions. As stated by Maddox and Ashby (1993; see also Maddox & Ashby, 1998), these are the simplest perceptual representation assumptions allowed in decision bound theory and are surely incorrect in most cases. We made this assumption for two reasons. First, this simplifies the modeling procedure because only one distribution parameter is free to vary. Because of the limited number of degrees of freedom in the accuracy data, this was necessary. Second, such a simple perceptual representation places the burden of prediction on the decision bound.

Because the general linear classifier is "nested" within the general quadratic classifier (i.e., the general linear classifier is a special case of the general quadratic classifier in which the quadratic terms equal zero), likelihood ratio tests can be used to determine which model provides the "most parsimonious account of the data" (e.g., Ashby, 1992b; Wickens, 1982). The basic idea is to determine whether the potentially nonzero quadratic terms in the general quadratic classifier provide a statistically "significant" improvement in fit over the general linear classifier (whose quadratic terms equal zero).

Table A1 displays the goodness-of-fit ( $-\ln L$ ) values for each model by observer and experiment. The goodness-of-fit value for the most parsimonious model is in bold type. Interestingly, these analyses suggest that many observers utilized a quadratic boundary. In fact, all 4 observers in the linear integration experiment and 1 of the 4 observers in the selective attention experiment appeared to use a nonlinear category boundary. The remaining 3 observers in the selective attention experiment used a linear category boundary.

At first glance, these findings seem to undermine our ability to draw inferences about the RT-distance hypothesis from these data, because the observers did not appear to use the experimenter-defined category boundary. However, there are at least two reasons to believe that these data will still serve our purpose. First, and most importantly, there is evidence that the extremely sim-

ple perceptual representation assumptions made by the models has an impact on the form of the best-fitting category boundary. For example, consider a situation in which the perceptual representation for a set of stimuli violates the equal-variance assumption, and the observer uses a linear category boundary. Suppose one were to apply the general linear classifier and general quadratic classifier to the data under the assumption that the perceptual representation satisfies the equal-variance assumption. This is a situation in which the perceptual representation assumption for the models is incorrect. Maddox (in press; see also Maddox & Ashby, 1998) investigated (through Monte Carlo simulation) several situations of this sort and found that the general quadratic classifier often provides the most parsimonious account of the data, even though the observer's category boundary is actually linear. Maddox argued that the extra boundary parameters in the general quadratic classifier allow the model to ab-

sorb (or to account for) some of the error created by the incorrect perceptual representation assumptions. This work is relevant because it is very likely that the equal-variance assumption is violated in the present data. (Of course, it is also possible that other assumptions are violated, such as the assumption of normally distributed perceptual effects.) Thus, the fact that the general quadratic classifier provided the most parsimonious account of much of the present data does not imply that the observers used a quadratic category boundary. Second, the goal of the present study was to examine the effect of distance to the category boundary on RT. This requires only that the observer's category boundary preserve the distance-to-bound relations outlined in the text. An examination of Figures 3 and 7 suggests that the observer's boundary was similar to the experimenter-defined boundary—in general, preserving the distance-to-bound relations.

**Table A1**  
**Goodness-of-Fit ( $-\ln L$ ) Values for the General**  
**Linear Classifier (GLC) and the General Quadratic**  
**Classifier (GQC) by Experiment and Observer**

Observer	Selective Attention Experiment		Linear Integration Experiment		Nonlinear Integration Experiment
	GLC	GQC	GLC	GQC	GQC
1	<b>110.20</b>	109.81	199.72	<b>144.75</b>	249.67
2	233.79	<b>220.36</b>	158.95	<b>124.91</b>	284.76
3	<b>117.64</b>	115.34	116.53	<b>84.45</b>	112.68
4	<b>93.07</b>	90.14	142.58	<b>123.38</b>	n.a.

Note—Values for the “most parsimonious” model are presented in bold type. The GLC fits were excluded from the nonlinear integration experiment because the fits were extremely poor.

(Manuscript received July 1, 1996;  
revision accepted for publication April 23, 1997.)