Psychonomic Bulletin & Review
2003, 10 (1), 141-148

Decision-bound theory and the influence of familiarity

TOM VERGUTS, GERT STORMS, and FRANCIS TUERLINCKX
University of Leuven, Leuven, Belgium

In this article, we derive a nonparametric prediction from decision-bound theory (DBT). The crucial aspect that is tested is whether or not familiarity of a stimulus affects response time in categorization. We show that, for our design, DBT, extended with some reasonable and testable assumptions, predicts no familiarity effect. Our prediction is nonparametric in that, rather than fit a specific instantiation of general DBT, we posit only some general assumptions of this theory and derive the prediction from these assumptions. It is found that familiarity did have a strong impact on response time for at least half of our participants. We suggest that DBT is in itself incomplete and should be extended to account for the full range of available data.

Decision-bound theory (DBT) has been very successful in describing human categorization performance (see, e.g., Ashby & Maddox, 1992; DBT is also sometimes called general recognition theory: see, e.g., Ashby & Perrin, 1988; Ashby & Townsend, 1986). The theory makes predictions on both categorization accuracy (Ashby & Gott, 1988) and response time (RT; Ashby, Boynton, & Lee, 1994; Ashby & Maddox, 1994; Maddox & Ashby, 1996; Maddox, Ashby, & Gottlob, 1998). Furthermore, DBT presents a challenge for competing (and also popular) exemplar models of categorization (see, e.g., Kruschke, 1992; Nosofsky & Palmeri, 1997b).

We will focus here on the situation in which there are two categories (A and B) and in which stimuli are characterized on two dimensions. In this case, DBT assumes that people draw a *decision bound* in the two-dimensional space. This bound can be characterized by a function h, such that the bound consists of all points for which $h = 0$. This is also often called a *discriminant function.* Stimuli are assigned to Category A if $h < 0$ and to Category B if $h > 0$. The undecided case ($h = 0$) can be simply resolved by guessing (see, e.g., Maddox & Ashby, 1993).

This is the DBT in its most elementary form. Additional assumptions are needed, however, to make it testable. First, a stimulus is characterized not in physical space, but, rather, in perceptual space for the categorization process. The decision-bound h is placed in this perceptual space (see, e.g., Maddox & Ashby, 1993). A second essential assumption of DBT is that noise is added to the perceptual representation to account for the fact that sensory channels are noisy (Maddox & Ashby, 1993). Third, it is assumed that the decision bound that is actually used by the participant has the same functional form as the optimal decision bound (Maddox & Ashby, 1993). Fourth, it is assumed that RT is inversely monotonically related to the distance between the stimulus (representation) and the decision bound (this is called the *RT–distance hypothesis*; Ashby et al., 1994).

Although this set of assumptions entails a definite theory about human categorization, the theory is still quite general. This is true mainly because, first, the mapping from physical to perceptual space is unspecified; second, the shape of the decision bound is unspecified except for the assumption that it has the same functional form as the optimal bound; and third, only an ordinal relation is specified between RT and the distance from the stimulus (representation) to the decision bound. Because of these unspecified aspects, results arguing against DBT could be dismissed on the ground that specific (parametric) assumptions were made to test the theory. For example, Ashby et al. (1994) cited a paper by Hyman and Frost (1975) in which evidence was presented against the RT–distance hypothesis. However, as Ashby et al. (1994) note, the result of Hyman and Frost is not immediately relevant, because Hyman and Frost used a specific functional form relating distance and RT (specifically, an exponentially decreasing relation). In a similar vein, results presented by Nosofsky and Palmeri (1997a) that question DBT could also be dismissed, because specific instantiations of DBT were used.

Ashby et al. (1994) advocated a nonparametric approach to testing DBT. That is, rather than make specific assumptions, they derived properties from the general theory and investigated their empirical validity. This approach was successfully applied. For example, the authors found negative rank correlations between distance and RT, in accord with DBT. Note, however, that the only aspect that was nonparametric in their study was the relation between RT and distance: They did estimate the parameters

This project was partly supported by Grant G.0266.02 from the Science Foundation Flanders to G. S. and Paul De Boeck. We thank Koen Lamberts and Robert Nosofsky for their comments on this manuscript. The first author is now at Ghent University. Correspondence can be sent to T. Verguts, Department of Experimental Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent, Belgium (e-mail: tom.verguts@rug.ac.be).

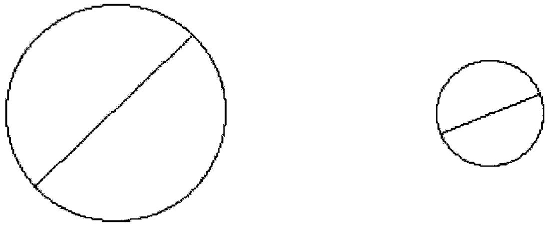

Figure 1. Example stimuli.

of a specific functional form of the bound h (the discriminant bound). They also found evidence against exemplar theory, in that familiarity of the stimulus showed no correlation with RT. Their notion of familiarity for a particular stimulus was the sum of its similarities to all other stimuli. In the present study, we will investigate the role of another measure of familiarity—namely, the number of times that the stimulus was presented. In the context of the development of automaticity, the effect of frequency of presentation of a stimulus has already been investigated by Lassaline and Logan (1993) and Palmeri (1997). In these studies, frequency of presentation was not manipulated while the distance to the boundary was held constant. However, this manipulation is critical to obtaining a strong test of DBT. Such a manipulation was performed by Nosofsky and Palmeri (1997b, Experiment 2), who found an effect of presentation frequency even if the distance to the decision bound was held constant. Nosofsky (1991) found a similar effect of presentation frequency. Our approach is different from his, however, in that we use the randomization technique favored by DBT theorists (see, e.g., Ashby & Gott, 1988): All analyses are performed at the level of the individual participant. Also, crucially, no parametric assumptions are imposed on DBT in order to derive predictions from it. The next section describes our approach in greater detail.

A PREDICTION FROM DECISION-BOUND THEORY

Suppose that stimuli are characterized on two dimensions—for instance, that they are circles of varying sizes, each containing a radial line of varying orientation (as in Shepard's, 1964, seminal study; see also Ashby et al., 1994, Experiment 3). Two examples are shown in Figure 1. In the stimulus space, a parallelogram is drawn, and stimuli are sampled from a distribution over the interior and boundary of this parallelogram, after the randomization technique introduced by Lee and Janke (1964). In Figure 2, such a parallelogram is depicted with corners x_a, x_b, x_c, and x_d. A linear bound separates the stimuli of Categories A and B (denoted h_{opt} in Figure 2); it is the line that crosses two of the parallel sides of the parallelogram at midline.

In the following, suppose that the mapping from physical to perceptual space yields no strong deformations. This assumption was also made by Ashby et al. (1994) for the range of stimulus values that was used in their (and our) experiment. The validity of this assumption will be discussed later. Still, the first dimension is measured in centimeters and the second in angular degrees, so the units of the two dimensions are not directly comparable, in that equal displacements along both dimensions may not correspond to similar psychological displacements (Shepard, 1964). Therefore, let us allow for the fact that one or both of the dimensions can be linearly rescaled to make the (psychological) units of the two dimensions equal. Since a (dimension-wise) linear scaling of a parallelogram yields another parallelogram, the perceptual space and the optimal boundary in it (h_{opt}) are as depicted in Figure 2. The four corners of this parallelogram will turn out to be critical in our analysis, and they are denoted a, b, c, and d, with coordinates x_a, x_b, x_c, and x_d, respectively.

As can be seen, stimuli to the right of h_{opt} belong to Category A, whereas stimuli to the left of h_{opt} belong to Category B. Suppose that $h = h_{\text{opt}}$—that is, that the participant uses the optimal decision boundary. Let T_a denote the RT on a presentation of Stimulus a, let $E(T_a)$ be its expectation, and let similar definitions apply to Stimuli b, c, and d. Then, if all the stimuli have the same perceptual noise distribution, it follows that DBT predicts that $E(T_a) = E(T_b) = E(T_c) = E(T_d)$. This is because the (orthogonal) distances of the four stimuli to the decision bound are all equal, which can be proven using elementary geometry. The result holds without specific assumptions on the RT–distance functional relation. However, we have now assumed that the actual bound (h) equals the optimal bound (h_{opt}). A more general result is needed to obtain a conclusive test of DBT. To obtain this, we specify the following four assumptions:

1. The mapping from physical to psychological space yields no strong deformations (in particular, only linear and dimension-wise transformations occur).
2. Noise for each stimulus follows a symmetric distribution, which is identical for all stimuli.
3. There is no criterion noise that is, once the percept of the stimulus is given, a response is determined by the side of the boundary at which the percept resides.
4. The actual bound (h) is parallel to the optimal bound (h_{opt}).

Evidence for the validity of each of these assumptions will be discussed later. The first three assumptions conform to assumptions commonly made in other DBT studies (e.g., Ashby et al., 1994; Maddox & Ashby, 1993; Maddox, Ashby, & Gottlob, 1998). The fourth is not common, but evidence was gathered in favor of it, as will be discussed later. Also, it can be replaced by the assumption that there is no perceptual bias, by which we mean that the bound h cuts the parallelogram in such a way that each of Categories A and B corresponds to equally sized areas of the parallelogram. One special case of a bound without perceptual bias would be if $h = h_{\text{opt}}$. However, this alternative assumption will not be discussed further.

The result we will use can now be stated as follows: If the probability of choosing the correct category (A or B)

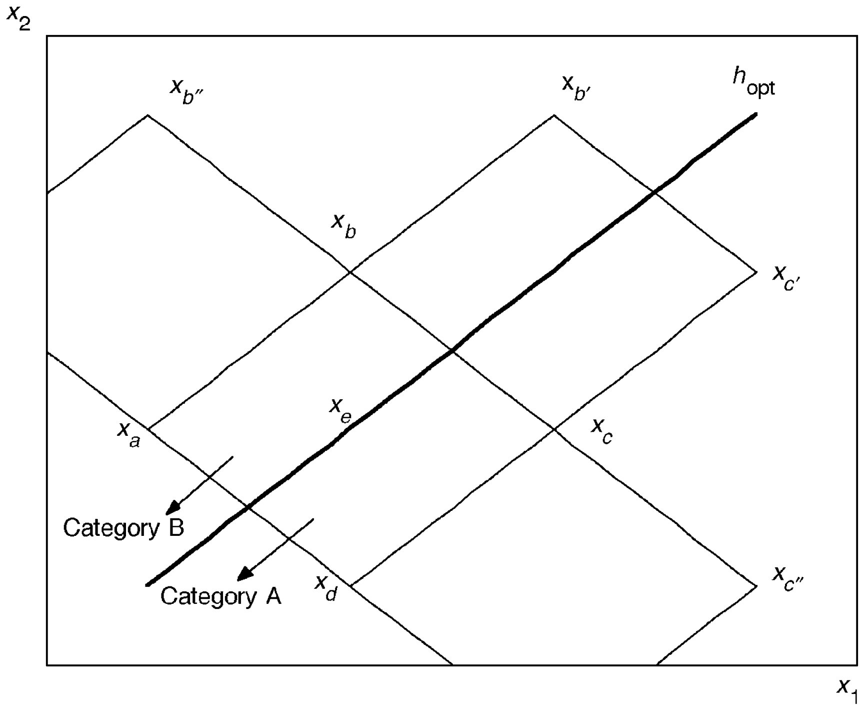

Figure 2. Parallelogram depicted in perceptual space.

is larger than .50 for each of the four Stimuli *a*, *b*, *c*, and *d* (i.e., if participants succeed in categorizing at better than chance level) and the decision bound *h* is parallel to the optimal bound, then the contrast $\psi = [E(T_b) + E(T_c)] - [E(T_a) + E(T_d)] = 0$.

To rephrase this, we note that if the probability of success is above chance for these four stimuli, the bound *h* must assign Stimuli *a* and *b* to Category B and Stimuli *c* and *d* to Category A. Next, for all such (linear) bounds that are parallel to h_{opt}, it is possible to prove that $\psi = 0$.[1] The intuition is that under the assumptions presented above, the distances of the perceptual representations of Stimuli *a* and *b* to the decision bound are equal and, therefore, $E(T_a) = E(T_b)$. Furthermore, the distances of the perceptual representations of Stimuli *c* and *d* to the decision bound are also equal, so $E(T_c) = E(T_d)$, and these two equalities together imply that the contrast $\psi = 0$. The variable ψ is a simple contrast, and the statistical significance of its estimate ψ is easily tested.

Suppose now that Stimuli *a* and *d* are shown more frequently than Stimuli *b* and *c*. According to DBT, the result still holds in this case. On the other hand, a model that incorporates frequency effects (see, e.g., Nosofsky & Palmeri, 1997b) would predict that $\psi > 0$. Although such models are not formally described, let alone tested in this article, this statement makes sense, because *a* and *d* can be assumed to have a higher representational strength (for a specific implementation of this idea, see Nosofsky & Palmeri, 1997b). One can then also expect that the RT to Stimuli *a* and *d* will be shorter than that to Stimuli *b* and *c*, and so $\psi > 0$. Indeed, such frequency effects were shown to be well captured by the exemplar-based random-walk model proposed by Nosofsky and Palmeri (1997b). The contrast ψ will be investigated in the experiment described in the following section.

The validity of our conclusion depends, of course, on the validity of the four assumptions. So, does it make sense to posit these assumptions at all? After the main results are discussed, we will elaborate on their validity and discuss what happens if these assumptions are violated.

METHOD

Participants

Eight students at the University of Leuven were recruited to participate in the experiment. They were paid for their participation.

Materials

The stimuli were circles with radial lines, such as those shown in Figure 1. The radii varied from 2.5 to 6 cm, and the orientations varied from 26º to 65º.[2]

The stimuli were sampled from a uniform distribution over the entire area of the parallelogram with corners at (2.5, 45), (4.25, 65), (6, 45), and (4.25, 26). These four corners are denoted Stimuli *a*, *b*, *c*, and *d*, respectively (see Figure 2). The first coordinate is the radius (in centimeters) of the circle, and the second coordinate is the orientation (in angular degrees) with respect to the horizontal line. In the following, the term *stimulus* will be used to identify an object with a fixed set of coordinates, whereas the term *item* will be used to identify a position in the testing sequence (e.g., "Item 17").

Table 1
Accuracies on the High-Frequency (HF) and Low-Frequency (LF) Corner Stimuli

	Participant							
Stimulus	1	2	3	4	5	6	7	8
First HF stimulus (*a*)	.78	.98	1	1	.97	1	.97	1
Second HF stimulus (*b*)	.67	.50	.83	1	1	.83	1	1
First LF stimulus (*c*)	.67	1	1	1	1	1	1	1
Second LF stimulus (*d*)	.72	.98	.98	.90	1	1	1	1

Design

Four hundred items were presented in blocks of 10. Of these, 20% were stimulus *a*, 20% were stimulus *d*, 2% were stimulus *c*, and another 2% were stimulus *b*. The remaining 56% consisted of stimuli from the uniform distribution described in the previous paragraph. As a way of counterbalancing, stimuli *b* and *c*, rather than stimuli *a* and *d*, were the frequent stimuli for half of our participants (e.g., *b* was presented 20% of the time and *a*, 2% of the time). However, this counterbalancing turned out to have no effect, so we will ignore it in what follows. We will refer to the *high-frequency* (HF) stimuli, which are stimuli *a* and *d* in Condition 1, and to the *low-frequency* (LF) stimuli, which are stimuli *b* and *c* in Condition 1. In Condition 2, the correspondence is exactly reversed. It is understood that the relevant contrast ψ is $[E(T_a) + E(T_d)] - [E(T_b) + E(T_c)]$ in Condition 2, so the contrast to be discussed is always $\psi = [E(T_{\mathrm{LF1}}) + E(T_{\mathrm{LF2}})] - [E(T_{\mathrm{HF1}}) + E(T_{\mathrm{HF2}})]$, where LF1 denotes the first LF stimulus, LF2 the second LF stimulus, and so on.

The category border (which is also the optimal decision bound) was as presented in Figure 2. More precisely, the category border crossed the *ad* and *bc* boundaries at midline. All stimuli to the right of this boundary were in Category A, and all stimuli to the left were in Category B. Hence, if there were no noise, use of this optimal bound would result in 100% accuracy.

After the first 400 items, a final block of extreme stimuli was shown. Specifically, consider a parallelogram constructed by copying the existing parallelogram and placing it to the side of the *bc* boundary (see Figure 2). This new parallelogram has corners *b*, *c*, *b′*, and *c′*. The latter two corners were presented as Items 401 and 402, respectively. Furthermore, imagine two extra copies of the parallelogram placed alongside *ab* and *cd*, respectively. Their corners on the line *bc*, in addition to *b* and *c* (*b″* and *c″*), were presented as Items 403 and 404, respectively. Their most extreme coordinate values are 2.5 and 7.75 cm for the radius dimension (for Stimuli *b″* and *c″*, respectively; see Figure 2), and 83º and 26º for the angle dimension (for stimuli *b″* and *c″*, respectively; see Figure 2). After these four stimuli were shown, a stimulus that was close to the optimal boundary was shown. Finally, this sequence of five stimuli was repeated. Hence, in total, 410 items were presented in 41 blocks (10 items in each block). These extreme stimuli are in line with the stimuli used by Nosofsky (1991), although in the present design they are only secondary to the main argument concerning the contrast ψ.

Procedure

Each participant was tested individually. The experimenter stayed in the experimentation room during the instructions (which were presented on the screen) to answer any possible questions. Stimulus presentation was response terminated.

RESULTS

Consider the contrast $\psi = [E(T_{\mathrm{LF1}}) + E(T_{\mathrm{LF2}})] - [E(T_{\mathrm{HF1}}) + E(T_{\mathrm{HF2}})]$. If its estimate ψ turns out to be statistically significant, there is evidence that DBT is violated. The contrast was calculated for each participant separately.

The first 100 items were not analyzed, because they served as practice trials. This is in line with the procedure followed by Ashby et al. (1994, p. 16). The reasoning was that exemplars may be used in that phase, according to DBT, but only to construct the decision bound. Hence, frequency effects in the practice items are not revealing.

First, we check the necessary condition that the accuracies on Stimuli *a*, *b*, *c*, and *d* be (relatively) high. Then the RT results of stimuli *a*, *b*, *c*, and *d* in Items 101 to 400 are discussed, followed by the RT results of the extreme stimuli shown in Items 401 to 410. Finally, we consider the possibility that one or more of the posited assumptions has been violated.

Accuracies

First, it has to be demonstrated that Stimuli *a*, *b*, *c*, and *d* were well learned in the sense that the probability of success was larger than .50. Table 1 shows the mean accuracies on these four stimuli for each participant. It can be seen that the four stimuli were indeed well learned: Only one participant had an accuracy of .50 on one stimulus. All other accuracies were higher, and almost 60% were perfect (accuracy of 1.0).

Response Times

In Items 101 to 400, the same percentages of Stimuli *a*, *b*, *c*, and *d* were used as in the complete set of the first 400

Table 2
Response Times of Each Participant for Each High- and Low- Frequency Stimulus and Corresponding Values of the *t* Statistic

Participant	First HF Stimulus	Second HF Stimulus	First LF Stimulus	Second LF Stimulus	ψ	*t* Test Results
1	747.5	883.4	999.7	741.3	110.1	$t(127) = 0.620, p = .269$
2	359.9	419.1	381.3	512.7	115.0	$t(128) = 2.327, p = .011$
3	625.0	774.8	831.7	678.5	110.4	$t(128) = 1.041, p = .150$
4	897.8	1,281.5	1,003.2	1,010.8	−165.3	$t(128) = -0.696, p = .756$
5	607.2	667.3	1,371.2	769.2	865.9	$t(120) = 5.342, p < .001$
6	707.1	759.9	1,248.2	840.2	621.4	$t(128) = 3.647, p < .001$
7	631.3	589.4	567.7	581.5	−71.5	$t(128) = -0.627, p = .734$
8	654.2	654.4	890.5	1,078.8	660.7	$t(128) = 3.493, p = .001$
All	653.750	753.725	911.688	776.625	280.8	$t = 15.147, p < .001$

Note—HF, high frequency; LF, low frequency. All *p*s are two-tailed.

items. The percentage of presentation of each of the HF stimuli was 20%, and for each of the LF stimuli it was 2%.

Table 2 presents the mean RTs for each of the HF and LF stimuli, the estimated ψ contrast, and the values for the t statistic corresponding to the ψ contrast for each participant separately, as was described in the previous section. A one-sided p value is reported as well. Under the null hypothesis of no differences between the four stimuli (i.e., $\psi = 0$), the t statistic should follow a t distribution with 128 degrees of freedom. However, Table 2 shows a deviating number of degrees of freedom for 2 participants. This is because an error was made in the number of stimuli of each type that was presented to these 2 participants: Participant 1 had one item fewer of type a, and Participant 5 had one, two, and five items fewer of types b, c, and d, respectively. However, the predictions remain the same in these two cases, and their data cause no problem in the analysis.

It can be seen that the t statistic tends to be high and does not seem to be centered around zero. For 4 participants, it was significant at level .05. By way of meta-analysis across participants, we added all the t statistics and compared the total to its (simulated) relevant distribution—a sum of t distributions. This resulted in a value of 15.147 (see bottom row of Table 2), which was 5.440 standard deviations above the (simulated) mean. Of 10,000 simulations, no value was higher than the observed one, so $p < .0001$.

Although Ashby et al. (1994) followed the same procedure of including items from Trial 101 onward only, it might be argued that learning was not yet completed after the first set of 100 items. Hence, the effect may be present only because the decision bound was still "under construction" and, hence, exemplars were still used after Block 1. Such an explanation predicts that there were differences in mean RT between Blocks 2, 3, and 4. Analyses of variance (ANOVAs) performed for each participant separately revealed that there were significant differences at level .05 between Blocks 2, 3, and 4 for 3 participants. However, 2 of them (Participants 4 and 7) contributed an opposite effect, so our conclusion is certainly not based on these 2 participants. For Participant 6, on the other hand, we did find strong differences among Blocks 2, 3, and 4, and Participant 6 contributed to the overall effect. After Block 2, learning seemed to be complete for this participant, because if only Blocks 3 and 4 were included in the ANOVA, the difference between blocks was no longer significant. We therefore calculated ψ for these two blocks only. The corresponding result was $t(88) = 3.597$, $p < .001$. In other words, it makes no difference whether Block 2 was included or not for this participant.

Although familiarity clearly had an overall effect, a relevant question is why the effect was nonsignificant for 2 participants and was even in the inverse direction for 2 others. One possibility is that there may be different baseline RTs for some of the stimuli and some participants. For example, some of the participants may have found the LF stimuli easier to encode than the HF stimuli, which may have caused a shorter baseline RT for the former than for the latter. In that case, this baseline effect would first need to be overcome before the practice effect could become noticeable. To investigate this, we calculated, for Participants 1, 3, 4, and 7 (those with a nonsignificant or inverse effect), the previously described ψ statistic, in which, for the HF stimuli, we took into account only the first eight stimulus presentations, to match the number of presentations for each of the LF stimuli. (Note that now the training items are used also, to control the number of previous presentations for each stimulus type.) If the HF stimuli had a higher baseline RT (i.e., longer RTs), a significant and negative effect ($t < 0$) should have been obtained. No effect was found for Participants 1, 3, or 4 (all ps $> .05$), but a strong effect was found for Participant 7 [$t(28) = -3.384$, $p = .001$, one-tailed]. This suggests that the reason the effect was in the inverse direction for this participant was the higher baseline difficulty of the HF stimuli for this participant.

Effects for the Extreme Stimuli

Next, we looked at the final block: Items 401 to 410. First, we considered Items 401 to 404, where Stimuli b', c', b'', and c'' were presented, respectively. At least one of Stimuli b' and c' is far from the decision bound (see Figure 2). Therefore, DBT predicts that the RT to at least one of these stimuli is small relative to the RT to stimuli in the interior of the parallelogram. Concerning the corner stimuli, DBT predicts equal mean RTs for a, b, c, d, b', and c' in case the decision boundary is optimal. However, if we exclude corner Stimuli a, b, c, and d, then DBT predicts, for at least one of Stimuli b' and c', a shorter RT than for the stimuli in the interior. Similarly, at least one of Stimuli b'' and c'' is further away from the decision bound than are the stimuli in the parallelogram, even if Stimuli a, b, c, and d are included. Therefore, extreme RTs could be expected here also.

To check this, we considered the cumulative RT distribution of each participant separately and investigated the percentile of Stimuli b', c', b'', and c'' (i.e., the percentage of RTs that were shorter than the RTs to these stimuli). When the distribution for Stimuli b' and c' was investigated, corner Stimuli a, b, c, and d were excluded. According to DBT, the resulting percentiles should be low.

However, this clearly was not the case of the data. The mean value (over participants) of the minimal percentile (over Stimuli b' and c') was equal to 52.375. For Stimuli b'' and c'', the mean value of the minimal percentile was 60.563.

It might be argued that the analyses in the previous paragraph are unfair to DBT, because the stimuli were very extreme relative to the previous ones and, hence, may have surprised the participants, resulting in a long RT. Although this is not an assumption of DBT, it might be argued that DBT will hold if this surprise effect is excluded. With this aim, we presented the extreme stimulus sequence anew as Items 406 to 409. If DBT is correct, short RTs are predicted on this (repeated) stimulus sequence, since the surprise effect can no longer be present. However, the data showed values very similar to those of the first presentation of this sequence. For Stimuli b' and c', the mean value was 53.713—virtually the same as for the first presentation. For Stimuli b'' and c'', the mean value was 55.425, which is also quite close to the value for the first presentation.

Validity of the Four Assumptions

We will now discuss how plausible it is to posit each of the four assumptions noted earlier. Each assumption is discussed consecutively.

Assumption 1. This is the assumption that no strong deformations occur from physical to perceptual space (specifically, only linear, dimension-wise transformations occur). The first dimension over which the stimuli vary is length of the radius of the circle. It is known that the Stevens exponent of length is very close to 1 (see, e.g., Maddox et al., 1998). This suggests that a linear rescaling assumption is warranted. Furthermore, the second dimension is the orientation of the radial line. It is known that for such stimuli, perceptual space is especially "stretched out" near the 0º and 90º orientations (i.e., horizontal and vertical lines; see, e.g., Vogels & Orban, 1986). However, our stimuli had radial lines with orientations ranging from about 26º to about 65º (see the Method section). Therefore, all of the stimuli had orientations between the horizontal and the vertical orientations, with the two extremes (26º and 65º) about equidistant from the horizontal (0º) and the vertical (90º) orientations, respectively. Even the extreme Stimuli b', c', b'', and c'' had orientations well between 0º and 90º, the minimum and maximum being 26º and 83º, respectively (see the Method section). In conclusion, the most nonlinear part of the orientation continuum was avoided in our stimuli. Even more relevant, Nosofsky (1985) performed multidimensional scaling (MDS) analyses on similar stimuli (circles of varying sizes containing radial lines with different orientations) and found the perceptual (MDS) space to be very similar to the physical space. These results together support the validity of this assumption.

Assumption 2. The second assumption concerns the noise distribution associated with each stimulus; it entails the fact that the noise distributions are symmetrical (e.g., normal) and equal for all stimuli. The part of the assumption concerning symmetry is quite weak; however, the equality for all stimuli may be questioned. Specifically, it might be assumed that the HF stimuli are more discriminable and, therefore, possess a smaller variance. To put it more concretely, let us assume that the noise distributions are normal, with a covariance matrix that is a scalar multiple of the identity matrix (which is also in line with standard assumptions of DBT theorists; e.g., Maddox et al., 1998). Then the HF stimuli may have a smaller scalar value associated with the covariance matrix. This implies that, with a finite sample size, the shortest RTs associated with the two LF stimuli are shorter than the shortest RTs associated with the HF stimuli. However, the present design does not allow us to investigate this issue in great detail, since the tails of the RT distribution are not well sampled. It remains for future research to investigate the plausibility of this assumption.

Assumption 3. The third assumption—no absence of criterion noise—is a technical assumption needed in the derivation of the central result stated earlier, and it is unclear how violation of this assumption would break down the central prediction that $\psi = 0$. First of all, the criterion noise does not influence RTs. The assumption of no criterion noise is needed only for the result that, if all four corner stimuli are assigned (on average) to the correct category, then the decision bound must assign each pair of stimuli of the same category to the same side. Furthermore, it seems that this result is robust to violation of Assumption 3, since the criterion noise follows a normal (and thus symmetric) distribution around zero.

Assumption 4. This is the assumption that the actual decision-bound h is parallel to the optimal decision-bound h_{opt}. To check this assumption, we fitted the decision-bound model to the accuracy data of all items from 101 to 400. Three parameters were estimated—namely, the intercept and slope of h and a noise variance parameter (as in Maddox & Ashby, 1993, p. 55). The analysis was performed for each participant separately. Furthermore, a restricted model was fitted, in which the slope was fixed at optimal, so the restricted model has only two free parameters. The relative fit of these two models gives an indication of the validity of the assumption of parallelism: Specifically, if the restricted (two-parameter) model does not a have a significantly lower fit value than the general (three-parameter) model, this indicates that Assumption 4 is plausible. Since the two models are nested, this can be assessed easily by constructing the likelihood-ratio statistic $-2*\ln(L_{Res} / L_{Full})$, where L_{Res} and L_{Full} denote the maximum likelihood for the restricted (two-parameter) model and the full (three-parameter) model, respectively.

For 7 of the 8 participants, the restricted model did not fit significantly worse than the full model did at level .05, suggesting that, if DBT holds, the bound of these participants is parallel to the optimal bound. For one participant, the restricted model fitted significantly worse [$\chi^2(1) = 13.910, p < .001$]. However, this was Participant 7, one of 2 participants with a negative contrast. If this participant is excluded from the meta-analysis reported above, we obtain an aggregate value that is 5.978 standard deviations above its theoretical mean ($p < .001$). This analysis suggests that the assumption of parallelism is plausible, and, to the extent that it is not plausible, the violation does not influence our conclusion.

DISCUSSION

Our design was very much in the spirit of earlier studies on the DBT of categorization by Ashby and his coworkers. First, we used the randomization technique proposed by Ashby and Gott (1988; see also Lee & Janke, 1964). Second, we used the stimuli (circles with radial lines), parameters (radius and orientation of angular line), and range of parameters used in Maddox and Ashby's (1993) Experiment 3. Third, like Ashby et al. (1994), we made no specifications as to the functional form relating distance and RT, and we even extended the nonparametric design by not completely specifying or estimating the parameters of the decision bound except to check one of the critical assumptions. Fourth, like Ashby et al. (1994), we presented 400 items (plus 10 extra extreme items), the first 100 of which served as practice and were not analyzed (also in accordance with Ashby et al., 1994). However, the critical difference in our study was that some stimuli were presented more frequently than others, whereas stimuli in the previously mentioned studies were typically shown only once.

Despite these similarities, our data speak against the general validity of DBT. This was shown by the fact that the ψ statistic was not centered around zero, indicating that RTs were in general longer on the less frequently presented stimuli. Furthermore, the percentiles of the extreme stimuli in the RT distributions were not low, indicating that the distance of the stimulus to the decision bound does not in itself determine RT.

On the other hand, for some of the participants, the effect (the contrast ψ) was not statistically significant, and for 2 participants its value was in the direction opposite that predicted by a familiarity hypothesis.

One possible concern with our method is that the frequency manipulation could have distorted the stimulus space, thus placing the HF stimuli farther apart and making DBT a viable theory after all. We feel this is implausible, because it would imply that people in different frequency conditions for a particular stimulus would perceive different line lengths for that stimulus, and this would be rather counterintuitive. However, future research should be conducted to investigate this possibility in more detail, to see whether it is empirically valid or not.

Another problem may be our restrictive use of perceptual noise distributions. We have assumed them to be equal for each stimulus, and it may be that, when different noise distributions are posited, DBT may capture the frequency effect after all. However, with normal noise distributions, and if the covariance matrix is a scalar of the identity matrix, this still implies that (with a finite sample size) the shortest RTs on the LF stimuli are shorter than the shortest RTs on the HF stimuli. Our design did not allow us to investigate this issue in detail. Furthermore, it may be that by a change in the noise distribution (e.g., making it nonsymmetric) or the introduction of nonzero covariances, DBT may capture the frequency effects. Again, future research should be carried out to investigate this possibility.

One possible conclusion could be that categorization models should consider the effect of presentation frequency (as, e.g., in the model proposed by Nosofsky & Palmeri, 1997b). In our opinion, however, to fully address these data, a fruitful approach would be to take into account individual differences in processing strategies. Rather than assume that every participant follows the same general model (possibly with different parameter settings), it may be useful to consider that different participants may conform to different models, or even that the same participant may conform to different models at different points in time. The latter line of research has become popular recently. Among others, Erickson and Kruschke (1998), Smith and Minda (2000), and Johansen and Palmeri (2001) describe models that are mixtures of more elementary (e.g., exemplar, or rule-based) models, in which different strategies may dominate over time. Also Ashby, Alfonso-Reese, Turken, and Waldron (1998) combined a rule-based and a DBT-based model in their neuropsychological theory of categorization. The RULEX model of Nosofsky, Palmeri, and McKinley (1994) incorporates the idea that different people may follow highly idiosyncratic categorization strategies, combining (idiosyncratic) rules with exemplar storage.

To sum up, we have argued that DBT in its present form is incomplete. A further investigation of the "mixture" models mentioned in the previous paragraph may more fully describe the rich pattern of human categorization data.

REFERENCES

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, **55**, 11-27.

Ashby, F. G., & Gott, R. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.

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. (1994). A response time theory of perceptual separability and perceptual integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423-466.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124-150.

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

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

Hyman, R., & Frost, N. H. (1975). Gradients and schema in pattern recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 630-654). London: Academic Press.

Johansen, M. K., & Palmeri, T. J. (2001). *Representational shifts in category learning*. Manuscript submitted for publication.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 561-581.

Lee, W., & Janke, M. (1964). Categorizing externally distributed stimulus samples for three continua. *Journal of Experimental Psychology*, **68**, 376-382.

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., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, **60**, 620-637.

Nosofsky, R. [M.] (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, **38**, 415-432.

Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, **19**, 131-150.

Nosofsky, R. M., & Palmeri, T. J. (1997a). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception & Psychophysics*, **59**, 1027-1048.

Nosofsky, R. M., & Palmeri, T. J. (1997b). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 324-354.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 3-27.

Vogels, R., & Orban, G. A. (1986). Decision factors affecting line orientation judgments in the method of single stimuli. *Perception & Psychophysics*, **40**, 74-84.

NOTES

1. A formal proof of these two statements is available from the first author.

2. This range covers about 95% of the stimuli used by Ashby et al. (1994) in their Experiment 3.

(Manuscript received June 6, 2001;
revision accepted for publication March 6, 2002.)