

## Tests of the ratio rule in categorization

A.J. Wills, Stian Reimers, Neil Stewart, Mark Suret, and I.P.L. McLaren  
*Cambridge University, Cambridge, U.K.*

Many theories of learning and memory (e.g., connectionist, associative, rational, exemplar based) produce *psychological magnitude terms* as output (i.e., numbers representing the momentary level of some subjective property). Many theories assume that these numbers may be translated into choice probabilities via the ratio rule, also known as the choice axiom (Luce, 1959) or the constant-ratio rule (Clarke, 1957). We present two categorization experiments employing artificial, visual, prototype-structured stimuli constructed from 12 symbols positioned on a grid. The ratio rule is shown to be incorrect for these experiments, given the assumption that the magnitude terms for each category are univariate functions of the number of category-appropriate symbols contained in the presented stimulus. A connectionist winner-take-all model of categorical decision (Wills & McLaren, 1997) is shown to account for our data given the same assumption. The central feature underlying the success of this model is the assumption that categorical decisions are based on a Thurstonian choice process (Thurstone, 1927, Case V) whose noise distribution is not double exponential in form.

Many theories of learning and memory employ what we shall describe as *psychological magnitude terms*. A psychological magnitude term is a continuous number that represents the momentary level of some subjective property. For example, a magnitude term could represent the subjective level of evidence for the belief that a stimulus is a member of a particular category.

Theories that employ psychological magnitude terms include connectionist and associative models, many exemplar-based theories, and some normative models. Any connectionist model that relies on the activation of output representations to predict responding is employing magnitude terms (e.g., Gluck & Bower, 1988; Kruschke, 1992; McClelland & Rumelhart, 1985; Seidenberg & McClelland, 1989). Similarly, the end product of many associative learning models is a number (e.g., the sum of associative strengths of cues present on a trial) that represents a prediction about level of responding (see e.g., Pearce, 1987; Rescorla & Wagner, 1972; Wagner, 1981). Further, many theories

---

Requests for reprints should be sent to A.J. Wills, University of Exeter, School of Psychology, Washington Singer Laboratories, Perry Road, Exeter, EX4 4QG, U.K. Email: a.j.wills@exeter.ac.uk

The authors would like to thank Mike Aitken, Fergal Jones, Koen Lamberts, Donald Laming, Stephen Monsell, and Thomas Palmeri for their helpful comments. This research was supported by grants from the B.B.S.R.C. and the E.S.R.C. to I.P.L. McLaren.

that posit the storage of specific examples or instances are magnitude based. For example, models that calculate summed similarity to stored instances in order to predict responding are magnitude based, with the summed similarity scores being the magnitude terms (e.g., Estes, 1994; Kruschke, 1992; Lamberts, in press; Medin & Schaffer, 1978; Nosofsky, 1986). Finally, some normative models—for example, Anderson's rational model (Anderson, 1991)—employ magnitude terms.

All models whose output is a set of magnitude terms must specify how these numbers translate into empirically testable predictions. In this paper we focus on a particular aspect of this *decision mechanism* issue—namely, how are predictions about the probability of a specific response derived when multiple responses are possible? An answer common to otherwise disparate memory models is that one should invoke the ratio rule. The ratio rule has been proposed by a number of authors (e.g., Bradley, 1954; Clarke, 1957; Luce, 1959) and goes under a variety of names, including the choice axiom (Luce, 1959) and the constant-ratio rule (Clarke, 1957). For our current purposes, the ratio rule may be expressed as

$$P(i) = \frac{v_i}{\sum_{j=1}^n v_j} \quad 1$$

where  $P(i)$  is the probability of choosing alternative  $i$  from  $n$  alternatives, and  $v_j$  is the magnitude term for the  $j$ th alternative. Theories that employ the ratio rule include McClelland and Rumelhart's model of word perception (McClelland & Rumelhart, 1981, Equation 7), their distributed memory model (McClelland & Rumelhart, 1985, p. 174–175), Gluck and Bower's model of category learning (Gluck & Bower, 1988, Equation 3), Kruschke's ALCOVE (Kruschke, 1992, Equation 3) and ADIT (Kruschke, 1996, Equation 10) models, and the generalized context model (Nosofsky, 1986, Equation 5). In some cases, the magnitude terms employed are transforms of the model's output, the most common transform being

$$v = e^{ko} \quad 2$$

where  $o$  is the model's output, and  $k$  is a scaling constant. Equation 1 implies that magnitude terms are non-negative because negative terms would lead to probabilities of less than zero. Amongst the advantages of the exponential transform of Equation 2 is that it ensures that magnitude terms are non-negative.

Although the ratio rule is commonly used to derive probability predictions from magnitude-based models, theorists are seldom explicit about their reasons for choosing this particular decision mechanism. Research designed specifically to test the ratio rule provides some support for it, but this support is by no means definitive. For example, Bradley (1954) and Hopkins (1954) tested the ratio rule in the context of pair-comparison experiments. In a pair-comparison experiment there are  $n$  stimuli. For each possible pair of stimuli, subjects are asked to choose one stimulus on the basis of some criterion (e.g., pick the sweeter of two flavours). These studies show that response probabilities in a number of pair-comparison experiments are explicable in terms of the ratio rule. The nature of the support provided by both studies is a demonstration that there is a set of

magnitude terms (one for each stimulus), which, when substituted into the ratio rule, produce predictions that do not differ significantly from the observed data.

There are at least three problems with taking these studies as good evidence in support of the ratio rule. First, no alternative theory is tested so there may be other formulations that would work equally well for these data (e.g., the difference rule previously investigated by Jones, Wills, & McLaren, 1998). Second, the pair-comparison experiments examined typically involve only a few stimuli, and hence there are almost as many free parameters as there are data points. For example, in one typical test there were four different stimuli and hence four free parameters. There are six possible pair comparisons of four stimuli and hence only six response probabilities to be predicted. It would perhaps be surprising if the ratio rule did *not* pass a goodness-of-fit test under such circumstances. The third problem is that support for the rule is based on a null result. Given that the sample size in all tests is small, the results may be due to a lack of power in the statistical test rather than the adequacy of the ratio rule.

Another area of research designed to test the ratio rule concentrates on its prediction that the ratio of any two choice probabilities is constant, irrespective of the total number of alternatives. This property of the ratio rule allows it to predict *full-set* choice probabilities from *sub-set* choice probabilities, and vice versa. For example, in an experiment by Clarke (1957) the full-set decision was to determine which syllable had just been presented in noise, given six alternative syllables to choose from. The sub-set decisions involved determining the syllables presented given just three of the original six alternatives. The full-set choice probabilities were used to derive predictions, via the ratio rule, for the sub-set choice probabilities. These predictions were then compared with the observed choice probabilities. Clarke concluded that the level of agreement between predictions and observations was good, with 95% of predictions deviating by less than 0.1 from the observed data. A number of other studies have come to similar conclusions for a variety of stimuli (e.g., Pollack & Decker, 1960). The problem in accepting such studies as good evidence in support of the ratio rule is that, as in the analysis of pair-comparison experiments, no alternative theory is considered.

One alternative to the ratio rule is to assume that people always choose the alternative with the largest subjective magnitude term. This may seem unlikely because the probability of choosing an alternative in a specific situation seems constrained to be one (if its magnitude term is the largest), or zero (if it is not). However, if subjective magnitude terms are subject to random variation across different occurrences of the same stimulus then probabilities other than one and zero can be predicted.

If the magnitude terms are assumed to have a Gaussian distribution then this alternative theory corresponds to Thurstone's (1927) theory of judgement, with our term psychological magnitude basically corresponding to Thurstone's term *discriminal process*. As has been noted previously (e.g., Luce, 1959, p. 56) the ratio rule and Thurstone's theory can often make very similar predictions. However, Yellott (1977) proved for situations involving three or more choices that the predictions of Thurstone's theory and the ratio rule can be equivalent if and only if the distributions employed in Thurstone's theory are double exponential. For a two-choice situation there are distributions other than the double exponential that allow equivalence (e.g., an exponential distribution).

Yellott's demonstrations were for Case V of Thurstone's theory, which assumes that all distributions have the same variance.

Burke and Zinnes (1965) examined data from three different pair-comparison experiments and concluded that Thurstone's theory (Case V) predicted response probabilities somewhat better than did the ratio rule. In contrast, Hohle (1966) concluded from his examination of a different set of pair-comparison data that the ratio rule was superior. Both papers rely on the quantitative difference between observed and predicted response probabilities to discriminate between the two theories—they calculate  $\chi^2$  goodness-of-fit for each theory and favour the theory that generally has the lower  $\chi^2$ . Taken together, it is unclear from these papers which theory is to be preferred.

In the current paper we report two experiments designed to test some predictions of the ratio rule for situations other than a straightforward two-choice task. The predictions of the ratio rule for these situations are unambiguous and amenable to experimental test. However, these predictions are qualitatively different from those made by a model based on the principles of Thurstone's theory. This alternative model is the winner-take-all (WTA) connectionist network proposed by Wills and McLaren (1997) and further investigated by Jones et al. (1998). Due to the length and complexity of exposition required, details of this model and the predictions it makes are presented towards the end of the paper.

## A prediction of the ratio rule

The prediction of the ratio rule under test in our experiments concerns a three-choice decision, a two-choice decision, and the relationship between the two. In the three-choice decision, the subject is presented with stimuli to which three responses (A, B, C) are potentially appropriate. The subject must make one of these responses to each stimulus. In the two-choice decision the same stimuli are presented but one of the responses (A) is disallowed by the experimenter. The subject must make one of the two remaining responses. This procedure is an example of the full-set vs. sub-set manipulation described earlier. The prediction under test depends on the assumption that the magnitude terms for allowed alternatives are determined by the stimulus itself and are not affected by whether a two-choice or three-choice decision is requested.

The prediction concerns the relationship between two measures. The first measure is the probability with which subjects make Response A in the three-choice decision. By the ratio rule as stated in Equation 1, this probability is predicted to be

$$P(A: A,B,C) = \frac{v_A}{v_A + v_B + v_C} \quad 3$$

where  $v_x$  is the magnitude term for response  $x$  produced by a given stimulus.

The second measure concerns the probability of making Response B to a given stimulus in the two-choice decision and the probability of making Response B to an equivalent stimulus in the three-choice decision. Specifically, the measure is the difference between the two-choice and three-choice probabilities, expressed as a proportion of the three-choice probability. Formally stated, this measure, which we will refer to as  $q$ , is

$$q = \frac{P(B: B,C) - P(B: A,B,C)}{P(B: A, B, C)} \quad 4$$

Under the ratio rule, and assuming type of decision does not affect magnitude terms,  $q$  is predicted to be

$$q = \frac{[v_B/(v_B + v_C)] - [v_B/(v_A + v_B + v_C)]}{v_B/(v_A + v_B + v_C)} \quad 5$$

which simplifies to

$$q = \frac{v_A}{v_B + v_C} \quad 6$$

Compare Equations 3 and 6. If  $v_A$  is constant, then any change in  $q$  or  $P(A: A,B,C)$  must be driven by a change in  $(v_B + v_C)$ . Further, if  $v_A$  is constant, *any given change in  $(v_B + v_C)$  must produce the same direction of change in  $q$  as it does in  $P(A: A,B,C)$* —for example, an increase in  $(v_B + v_C)$  must lead to a decrease in  $q$  and a decrease in  $P(A: A,B,C)$ . The experiments presented in this paper test this prediction in the context of a category learning experiment.

### A specific prediction about category learning

Category learning is the task of learning the correct category label for each of a set of presented stimuli. It has been the subject of a large number of studies, typically involving stimuli that are visual, novel, and abstract, and a choice of labels that is limited to a few pre-defined alternatives (e.g., Homa, Sterling, & Trepel, 1981; Hull, 1920; Nosofsky, 1986; Posner & Keele, 1968, amongst many others). Often the stimuli for a given category are all distortions of a base pattern, which is sometimes referred to as a *prototype*.

Many theorists have proposed magnitude-based models to account for the data from category-learning experiments (e.g., Estes, 1994; Gluck & Bower, 1988; Kruschke, 1992; McClelland & Rumelhart, 1985; Medin & Schaffer, 1978; Nosofsky, 1986), and a great many of these models (including all those just cited) employ the ratio rule to translate magnitude terms into response probability predictions. The widespread use of the ratio rule in models of category learning suggests that this might be an apposite procedure for testing its predictions.

In the current experiments, subjects were presented with labelled examples from each of three categories (A, B, and C) and were then asked to decide the category membership of each of a set of unlabelled *transfer* examples. This general methodology has been employed previously to investigate a variety of issues (e.g., Homa et al., 1981; Posner & Keele, 1968). In our experiments, half the subjects were asked to decide whether each transfer stimulus was an A, a B, or a C. The other half were asked, for an equivalent set of transfer stimuli, whether each stimulus was a B or a C.

The stimuli employed were composed of a fixed number of distinct features. All transfer stimuli contained four Category A features and a varying number of Category B and Category C features. The number of Category C features was constrained by the equation

$$x_C = 8 - x_B$$

where  $x_C$  is the number of Category C features and  $x_B$  is the number of Category B features.

Subjects' responses to these transfer stimuli were used to estimate *generalization functions*. The generalization functions considered were the probability of making a particular response to a stimulus as a function of its similarity to a particular category prototype. Our index of similarity to a category prototype was the number of features from that prototype that the stimulus contained.

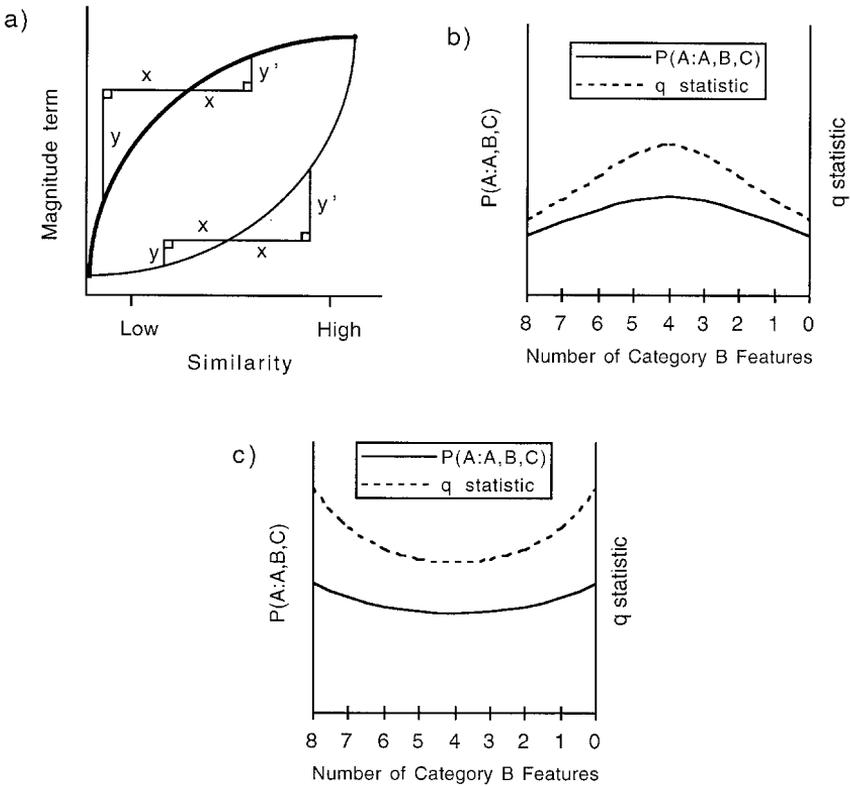
The specific generalization functions considered were  $P(A: A,B,C)$ ,  $P(B: A,B,C)$ , and  $P(B: B,C)$ , as a function of similarity to the Category C prototype. The value of  $q$  as a function of similarity to Category C can be calculated from the  $P(B: A,B,C)$  and  $P(B: B,C)$  functions. Hence, the experiment was designed to provide an assay of how  $P(A: A,B,C)$  and  $q$  change as a function of similarity to Prototype C (on our similarity index of number of Prototype C features). What one can conclude from these empirical functions depends on how the manipulation of number of A, B, and C features translates into changes in the magnitude terms for Responses A, B, and C. The crucial assumption we make is that the magnitude terms for a response are a *univariate* function of the number of corresponding features in a stimulus. In other words, we assume that the mean magnitude term for response  $y$  is solely determinable from the number of category  $y$  features in the presented stimulus.

Under this assumption, our experiment is a test of the prediction of the ratio rule derived in the previous section. The term  $v_A$  will be a constant because number of Category A features is constant. Therefore, any change in the observed values of  $P(A: A,B,C)$  and  $q$  as a function of number of Category C features must be caused by changes in  $(v_B + v_C)$ . Further, a given change in number of Category C features must result in the same change in  $(v_B + v_C)$  for the  $q$  function as for the  $P(A: A,B,C)$  function because the same stimuli are being considered in both cases. The ratio rule therefore predicts that the observed  $q$  and  $P(A: A,B,C)$  functions will show the same direction of change over any interval of our "number of Category C features" index. If our assumption of univariate magnitude functions holds, the accuracy with which this index represents psychological similarity as assessed by, for example, multi-dimensional scaling of identification responses (see e.g., Shepard, 1957), is not important.

### Further specific tests

In the current experiments, subjects get an equal amount of training on each of the three categories, and the method of stimulus construction is identical for each category. Hence, it might be reasonable to assume that each category response has the *same* univariate magnitude function: in other words, that the mean value for the magnitude term for response  $y$  is the same for a stimulus containing  $x$  category-appropriate elements, irrespective of whether  $y$  is A, B, or C. If one makes this assumption for Responses B and C then further predictions can be derived from the ratio rule. First,  $q$  and  $P(A: A,B,C)$  must be symmetrical around  $x_B = x_C$  due to the relation between  $x_B$  and  $x_C$  given in Equation 7. Second, the shape of the  $q$  and  $P(A: A,B,C)$  functions is determined by the shape of the magnitude function.

Previous categorization experiments have shown that the ratio rule can be rejected if magnitude is assumed to be a linear function of our similarity index, but not if it is assumed to be an exponential function (Jones et al., 1998). These experiments employed similar procedures and stimuli to those of our current experiments. If magnitude terms in the current experiments are an exponential (or any monotonically accelerating) function of our index then  $q$  and  $P(A: A,B,C)$  will be at a maximum where  $x_B = x_C$ . Where  $x_B = x_C$ , the magnitude terms for Categories B and C will be equal. Any departure from  $x_B = x_C$  will lead to an increase in one  $x$  term and a corresponding decrease in the other, due to the relationship given in Equation 7. This will result in an increase in one magnitude term and a decrease in the other. However, as the function is accelerating, the increase in magnitude produced by a given change in  $x$  will be greater than the decrease in magnitude produced by that change (Figure 1a may make this clearer: the thinner line is an accelerating function). This results in an increase in  $(v_B + v_C)$  and hence a decrease in  $q$  and  $P(A: A,B,C)$ , for any departure from  $x_B = x_C$ . This prediction is illustrated in Figure 1b.



**Figure 1.** (a) Two theoretical magnitude functions. The thinner line is a monotonically accelerating function. For any point on this curve, the change in magnitude caused by an increase  $x$  in similarity produces a larger change in magnitude ( $y'$ ) than does a corresponding decrease in similarity ( $y$ ). The converse is true for a monotonically decelerating function (thicker line). (b) Illustration of the ratio rule's predictions if Categories B and C have the same monotonically accelerating magnitude function. (c) Illustration of the ratio rule's predictions if Categories B and C have the same monotonically decelerating function.

Of course, the magnitude function may be neither linear nor monotonically accelerating. We consider one further class of magnitude functions in this paper—monotonically decelerating functions. If the magnitude terms for different categories are identical and of this form then  $q$  and  $P(A: A,B,C)$  will be at a minimum where  $x_B = x_C$ . As for accelerating decay functions, any departure from  $x_B = x_C$  will lead to an increase in one  $x$  term and a corresponding decrease in the other. However, because the function is decelerating, the decrease outweighs the increase (see Figure 1, thicker line), leading to a reduction in  $(v_B + v_C)$  and hence an increase in  $q$  and  $P(A: A,B,C)$ . This prediction is illustrated in Figure 1c.

## EXPERIMENT 1

To summarize, our first experiment had two phases. In the training phase, subjects learned about the category membership of training stimuli. Each training stimulus belonged to one of three categories—A, B, or C. In the generalization phase that followed, subjects were asked to determine the category membership of a different, transfer, stimulus set. Half of the subjects were allowed to respond A, B, or C. For the other half of subjects the response A was disallowed. In each group, the responses to transfer stimuli were used to derive generalization probability functions (the probability of making a specific response to a stimulus as a function of the difference between that stimulus and the appropriate category prototype).

The central prediction of the ratio rule applied to any magnitude-based model of categorization is that our measures  $P(A: A,B,C)$  and  $q$  derived from these generalization probability functions will show the same direction of change over any given section of the generalization function. This prediction is based on the assumption that the magnitude-based model of categorization produces magnitude terms for each category that are univariate functions of the number of category-appropriate features that the presented stimulus contains. Subsidiary predictions about the shape of the  $q$  and  $P(A: A,B,C)$  generalization functions, illustrated in Figure 1, can be derived by making further assumptions about the nature of the magnitude functions.

## Method

### Subjects and apparatus

The subjects were 24 adults, mainly Cambridge University undergraduate students. The experiment was run (by SR and NS) on two Acorn Risc PC computers in different, quiet cubicles. The computers were connected to 14" colour monitors (Acorn AKF 60), which were placed at eye level and about 80 cm in front of subjects. Responses were recorded via the "X", "B", and "M" keys of a standard PC keyboard. For the purposes of this experiment, the keys were re-labelled "A", "B", and "C" using bold red letters against a white background. This resulted in three response keys near the centre of the bottom row of the keyboard, separated from each other by one key and ordered "A", "B", "C" from left to right.

## Stimuli

Each stimulus was a collection of 12 different small pictures (hereafter *elements*) in a 4.5-cm by 3.5-cm rectangle outline, arranged on an invisible four-by-three grid (see Figure 2 for an example). Every stimulus contained 12 elements drawn from the pool of 36 that we have used in previous experiments (see Jones et al., 1998, p. 37; Wills & McLaren, 1997, p. 611). At the beginning of the experiment, and separately for each subject, 12 elements from the pool were randomly designated as Category A elements, a different 12 as Category B elements, and the remaining 12 as Category C elements. Each training stimulus for each category was constructed by starting with all 12 elements characteristic of that category (e.g., Category A elements for a Category A training stimulus). Each element in the training stimulus then underwent a 10% chance of being replaced by a randomly chosen element from one of the other two sets (e.g., replaced by a B or C element in the case of a Category A training stimulus). It was these modified stimuli that were presented to subjects as training stimuli. This procedure produces training examples that are composed predominately of elements characteristic of a particular category but that also exhibit considerable variability in terms of the specific elements they contain. A total of 90 training stimuli were created for each subject, 30 from each of the three categories.

Each test stimulus contained four A elements,  $x$  B elements, and  $(8 - x)$  C elements, where  $x$  could be 0, 1, 2, 3, 4, 5, 6, 7, or 8. Ten examples of each of these nine types of test stimulus were created for each subject. The specific elements used to create each test stimulus were chosen randomly within the constraints provided by the number of A, B, and C elements that the stimulus was to contain. Ten examples of each of four dummy stimuli were also created, these stimuli being (8 A, 0 B, 4 C), (8 A, 4 B, 0 C), (0 A, 4 B, 8 C), and (0 A, 8 B, 4 C). The purpose of the dummy stimuli was to obscure from the subjects that all test stimuli of interest (from the perspective of the experimenters) were constant in terms of the number of elements from Category A that they contained.

The position of elements within a stimulus was randomly determined for each stimulus presented, with the constraint that exactly one element occurred at each location in the four-by-three grid. Where stimuli were accompanied by a category label, this was presented as a large sans serif capital A, B, or C in an outline rectangle ( $4.5 \times 3.5$  cm) immediately to the right of the stimulus itself.

## Procedure

Subjects were allocated to one of two between-subject groups such that an equal number (12) participated in each. The two groups, referred to hereafter as the *two-choice* and *three-choice* groups, differed only in the question they were asked in the test phase.

After subjects had read some general instructions, the 90 training stimuli were presented to them sequentially and in a random order. Each example was presented for 5 s in the centre of the monitor, accompanied by the appropriate category label. A plain mid-grey mask in the stimulus and label rectangles preceded the next example for a period of 2 s. Subjects were not required to respond in any

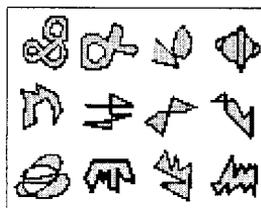


Figure 2. An example stimulus.

way in this first phase of the experiment. They were simply asked to concentrate on the examples shown as they would later be asked to classify new, unlabelled examples. This training procedure has proved effective for stimuli of this type in a number of previous experiments (Jones et al., 1998; Wills & McLaren, 1997).

The training phase was followed by a test phase. There were 130 stimuli in the test phase (90 target stimuli and 40 dummy stimuli), which again were presented sequentially and in a random order. Test stimuli were not accompanied by a category label. Subjects in the two-choice condition were asked, for each stimulus, "Is this a B or a C?" Subjects in the three-choice condition were asked "Is this an A, a B, or a C?". In both conditions, subjects responded by pressing the appropriate key on the computer keyboard. Subjects then pressed the "Y" key, whereupon the next stimulus was immediately presented. There was no time limit for these decisions, and subjects were put under no pressure to respond quickly.

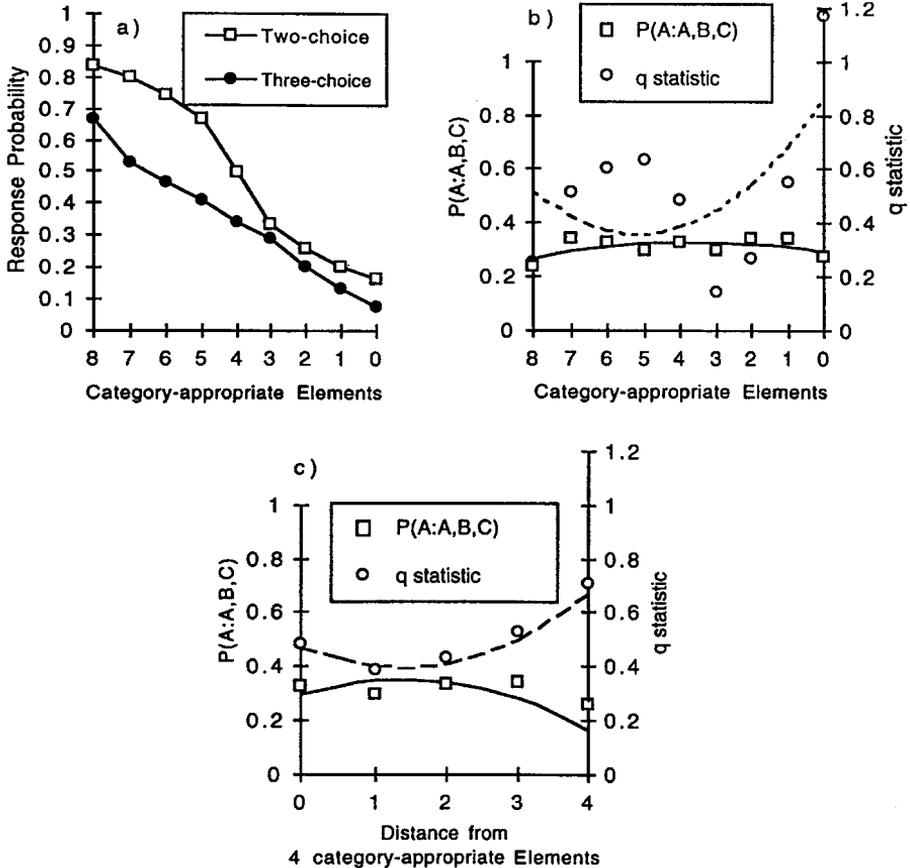
## Results

The probabilities with which subjects responded "B" and "C" to the test stimuli allow a test of our subsidiary assumption that the magnitude functions for Categories B and C are the same. If they are then the empirical generalization functions of the two categories should be the same. That is to say, the probability of responding "B" to a stimulus with  $x$  B elements should be the same as the probability of responding "C" to a stimulus with  $x$  C elements.

To assess whether this assumption could be disconfirmed with the current data set, type of probability assessment P(B) or P(C) was included as a within-subjects factor in a mixed-design analysis of variance (ANOVA). The other two factors in this analysis were: (1) the number of category-appropriate elements in a stimulus (i.e., B elements for P(B) and C elements for P(C), nine levels, within subject) and (2) experimental condition (two-choice vs. three-choice, two levels, between subjects). Type of probability assessment was non-significant as a main effect,  $F(1, 22) < 1$ ,  $p > .5$ , and did not interact significantly with any other factor,  $p > .05$  in all cases. However, this analysis did reveal that the type of question asked (two-choice or three-choice) had a significant effect on responding,  $F(1, 22) = 88$ ,  $p < .0005$ , as did the number of category-appropriate elements in a stimulus,  $F(8, 176) = 48$ ,  $p < .05$ , after a conservative correction for non-sphericity (Greenhouse & Geisser, 1959). The interaction between number of category-appropriate elements and experimental condition was non-significant,  $F(8, 176) = 2.1$ ,  $p = .13$ , after a Greenhouse-Geisser correction ( $\hat{\epsilon} = 0.27$ ).

As the two probability assessments did not differ significantly, their average was employed in all subsequent analyses. Figure 3a shows these means plotted as a function of number of category-appropriate elements for both the two-choice and the three-choice conditions.<sup>1</sup> Values for the  $q$  statistic (as stated in Equation 4) were then calculated from each pair of points. Specifically, for each value of category-appropriate elements, the mean three-choice probability was subtracted from the mean two-choice probability, and the resulting number was divided by the mean three-choice probability. The resultant values

<sup>1</sup> Note that the abscissa in all these figures is reversed such that category-appropriate elements reduce from left to right. This is to allow the figures to conform to the convention that generalization functions are plotted with negative slopes.



**Figure 3.** Results of Experiment 1. (a) Mean response probability as a function of number of category-appropriate elements in the presented test stimulus. The ordinate is the mean of the probability of producing a Category B response to stimuli containing a given number of Category B elements and the probability of producing a Category C response to stimuli containing that number of Category C elements. (b) Probability of producing a Category A response in the three-choice condition, and the  $q$  statistic (see Equation 4), as a function of number of category-appropriate elements. (c) As Figure 3b, but plotted as a function of “distance” from the four category-appropriate elements point on that graph. Figures 3b and 3c show data as unconnected plot symbols; the lines are the best-fitting quadratics to the plotted data. Figure 3a shows data as connected plot symbols; there are no best-fit lines shown.

of  $q$  are plotted as a function of category-appropriate elements in Figure 3b. Also plotted in Figure 3b is the mean probability of choosing Category A in the three-choice condition as a function of category-appropriate elements. This is the  $P(A:A,B,C)$  function referred to in the Introduction. Category B elements were used as the category-appropriate elements when plotting this function. Using Category C elements instead would simply reverse the function left to right.

The central prediction of the ratio rule under test is that the  $q$  and  $P(A:A,B,C)$  functions will show the same direction of change over any interval of the category-appropriate element axis. Inspection of Figure 3b suggests that is not the case for our

data. To investigate this further, we attempted to characterize the two functions by polynomial regression. Inspection of the  $P(A: A,B,C)$  function suggests that it exhibits a shallow inverted-U trend. Second-order polynomial regression of the nine data points of the mean  $P(A: A,B,C)$  function produced the best-fit line illustrated by the solid curve in Figure 3b. The equation of the line is  $P(A) = -0.003c^2 + 0.021c + 0.29$ , where  $c$  is the number of category-appropriate elements. A corresponding regression of the  $q$  function produced the best-fit line  $q = 0.02c^2 - 0.19c + 0.86$ , where  $c$  is the number of category-appropriate elements. This equation suggests a U-shaped trend in the  $q$  statistic, opposite to the trend in  $P(A: A,B,C)$  and hence contrary to the predictions of the ratio rule.

The preceding analysis is inconclusive, however, as neither equation represents a significant fit to the data,  $F(2, 6) = 1.6$ ,  $p > .25$ , for the  $P(A: A,B,C)$  function and  $F(2, 6) = 1.3$ ,  $p > .3$ , for the  $q$  function. Further inspection of the  $q$  statistic suggests that there may be a cubic component in the data. This issue is returned to in the Discussion.

Given that the generalization functions of Categories B and C are not significantly different to each other in the current data set, a further analysis of the  $q$  and  $P(A: A,B,C)$  functions is possible. Recall that the ratio rule predicts that if the magnitude functions for Categories B and C are the same then the both functions should be symmetrical about the point  $x_B = x_C$  (which is the point 4 category-appropriate elements). Although inspection of Figure 3b suggests that the  $q$  function is not symmetrical about this point, one might argue that its deviation from symmetry is due to sampling error. On the basis of this argument, one might average the pairs of data points at 0 and 8, 1 and 7, 2 and 6, and 3 and 5 category-appropriate elements on the grounds that these points were equivalent and hence taking the average should reduce sampling error. These pairs of points can be considered to have a "distance" from 4 category-appropriate elements of 1, 2, 3, and 4, respectively. The single data point at 4 category-appropriate elements provides the zero distance point.

Figure 3c shows  $q$  and  $P(A: A,B,C)$  as a function of distance from the 4 category-appropriate elements point. Second-order polynomial regression revealed that the  $q$  statistic was best fit by the function  $q = 0.04d^2 - 0.11d + 0.37$ , where  $d$  is the value on the ordinate of Figure 3c. This function, shown as a dotted line in Figure 3c, was a significant fit to the data,  $F(2, 2) = 116$ ,  $p < .01$ , and all three of its terms were significantly different from zero,  $t(3) = 9.8$ ,  $p < .005$  for the  $d^2$  term,  $t(3) = 6.1$ ,  $p < .001$  for the  $d$  term, and  $t(3) = 31$ ,  $p < .0005$  for the intercept. A corresponding regression performed on the  $P(A: A,B,C)$  data produced a best-fit line  $P(A: A,B,C) = -0.03d^2 + 0.08d + 0.30$ . This function, shown as a solid line in Figure 3c, suggests a downward trend but did not significantly fit the data,  $F(2, 2) = 2.9$ ,  $p > .25$ .

## Discussion

The results of Experiment 1 are somewhat at odds with the predictions of magnitude-based models of categorization that employ the ratio rule. Assuming magnitude is a univariate function of our category-appropriate elements index, the ratio rule predicts that the  $q$  and  $P(A: A,B,C)$  functions should show the same direction of change over any interval of that index.

The indications from Figure 3b are that this prediction is not upheld in Experiment 1. However, the failure of both the  $q$  and the  $P(A: A,B,C)$  functions to significantly fit a quadratic makes it difficult to draw firm conclusions. Quadratic functions were chosen because this is basically the form predicted by both monotonically accelerating and monotonically decelerating magnitude functions. One avenue of analysis would have been to attempt to fit higher-order functions to the data, such as a cubic function to the  $q$  statistic. We did, in fact, investigate some such functions, but the results of a replication of this experiment in Experiment 2 (reported later) subsequently led us to believe that the cubic component of the  $q$  statistic was not reliable. Hence, these analyses are not presented in this paper.

Some conclusions can be drawn from the current experiment under the additional assumption that the magnitude functions for categories B and C are the same. No evidence against this assumption was found in the direct test provided by the ANOVA reported earlier. Under this additional assumption, both the  $q$  and the  $P(A: A,B,C)$  functions should be symmetrical about the point 4 B elements. The fact that the observed  $q$  function is clearly not symmetrical about this point is potentially a problem for the ratio rule. Such a problem might be dismissed as sampling error by arguing that the function is symmetrical and that deviations from symmetry are due to the relatively small sample size. However, if such an appeal is valid, then it must be permissible to average data points at corresponding distances from the mid-point of 4 category-appropriate elements. The prediction of the ratio rule remains that the  $q$  and  $P(A: A,B,C)$  functions should show the same direction of change over any interval of the abscissa in Figure 3c.

Inspection of this figure strongly suggests that this prediction is not upheld. However, the failure of the  $P(A: A,B,C)$  function to fit any quadratic function whilst the  $q$  function fits an increasing function allows the possibility that both functions are, in fact, increasing but that the trend in  $P(A: A,B,C)$  is too shallow to be detected reliably. This interpretation of the results would be consistent with the application of the ratio rule to a magnitude-based model producing monotonically decelerating magnitude functions. However, the results of Experiment 1 are contrary to the predictions of the ratio rule when applied to a model producing linear or monotonically accelerating (e.g., exponential) magnitude functions. If the magnitude functions produced were monotonically accelerating then an inverted-U trend in both  $q$  and  $P(A: A,B,C)$  would be predicted by the ratio rule, and hence both functions in Figure 3c would be predicted to show a decreasing trend. If the magnitude functions produced were linear then both functions in Figure 3c would be predicted to be horizontal. The significant increasing trend in the  $q$  function is contrary to both these predictions. Hence, for models producing linear or monotonically accelerating magnitude functions, the ratio rule is an inappropriate theory of the decision process in categorization (within the assumptions made).

We felt that the results of Experiment 1 were sufficiently anomalous, from the perspective of the ratio rule theory, to merit replication. This was one of the purposes of Experiment 2.

## EXPERIMENT 2

One problem with the design of Experiment 1 was that category labels were not counter-balanced; it was always Category A that had the constant number of elements at test and that was the disallowed option in the two-choice condition. If our subjects had a consistent response bias towards one of the category labels then this might have led to a distortion of response probabilities in a way outside of the scope of the ratio rule as considered here. Experiment 2 controls for this by replicating Experiment 1, but employing equal numbers of subjects with A, B, and C as the category contributing a constant four elements in test stimuli and the category disallowed in the two-choice condition. For brevity, we shall refer to the category that provides a constant number of elements in test stimuli as Category a, and the two categories providing variable numbers of elements as Categories b and c.

Experiment 2 extends Experiment 1 by the addition of a third between-subjects condition. Subjects in this condition received the same training stimuli and test question as subjects in the three-choice condition. However, in this new condition the test stimuli contained no elements from Category a. Instead, the four Category a elements were replaced with four elements unseen in the training phase. This manipulation was designed to substantially reduce the probability with which subjects responded that test stimuli came from Category a, without actually disallowing this response option. The addition of this *novel-elements* condition allows the computation of two further functions. The first function we describe as  $P(b: a,b,c)'$ , which is the probability of choosing Category a in the novel-elements condition. The second function is  $q'$ , calculated as  $q$  but replacing  $P(b: a,b,c)$ , the mean probability from the two-choice condition, with  $P(b: a,b,c)'$ , the mean probability from the novel-elements condition.

These two new functions, along with the  $q$  function and the  $P(a: a,b,c)$  function from the three-choice condition, give a total of four functions to be assessed in Experiment 2. Our assumption was that our novel-elements manipulation would produce a constant, possibly zero, magnitude term for Category a at test (hereafter referred to as  $v_N$ ). Under this assumption, the ratio rule predicts that all four functions must show the same direction of change over any interval of our category-appropriate elements measure. The derivation of this prediction has already been demonstrated for  $q$  and  $P(a: a,b,c)$ . Its extension to the other two functions is straightforward. Predictions for the  $P(a: a,b,c)'$  function can be derived from Equation 3, simply by substituting  $v_N$  for  $v_A$ . The ratio rule's predictions for the  $q'$  function are

$$q' = \frac{P(b: a,b,c)' - P(b: a,b,c)}{P(b: a,b,c)} = \frac{[v_B/(v_N + v_B + v_C)] - [v_B/(v_A + v_B + v_C)]}{v_B/(v_A + v_B + v_C)} \quad 8$$

which simplifies to

$$q' = \frac{v_A - v_N}{v_B + v_C + v_N} \quad 9$$

Note that for both  $q'$  and  $P(a: a,b,c)'$  variability in the function is determined by the term  $(v_B + v_C)$  in the denominator, all other terms being constant. The same conditions hold for the  $q$  and  $P(a: a,b,c)$  functions. Hence, the ratio rule predicts that all four functions will show the same direction of change over any interval of category-appropriate elements.

## Method

### Subjects and apparatus

The subjects were 36 higher-education students from the Cambridge area; most were Cambridge undergraduate or graduate students. The experiment was run (by MS) on an Acorn Risc PC computer placed in one of two quiet experimental cubicles. The computer was connected to a 14" colour monitor (Acorn AKF 60), placed at eye level and approximately 90 cm in front of the subjects. Responses were recorded via a PC keyboard marked up in the manner described in Experiment 1.

### Stimuli

The pool of elements from which stimuli were constructed was extended from 36 to 40 elements (the 4 extra elements used are shown on the bottom row of Figure 2). The pool was extended so that once 12 elements had been randomly allocated to each of the three categories there would be 4 unallocated elements. These 4 elements were designated as the *novel elements*, and were selected independently for each subject. The novel elements were not used in the two-choice or three-choice condition. In the novel-elements condition, the 4 novel elements did not appear in training stimuli, but did appear in every test stimulus, along with  $x$  elements from Category b and  $(8-x)$  elements from Category c, where  $x$  took the values 0, 1, 2, 3, 4, 5, 6, 7, or 8. In other words, the test stimuli in the novel-element condition were constructed in the same manner as the test stimuli in the other two conditions, except that the 4 randomly selected elements from Category a were replaced by the 4 novel elements. In all other respects, the method of stimulus construction was identical to that employed in Experiment 1. As with all other elements, the position of the novel elements within a stimulus was determined randomly for each stimulus.

### Procedure

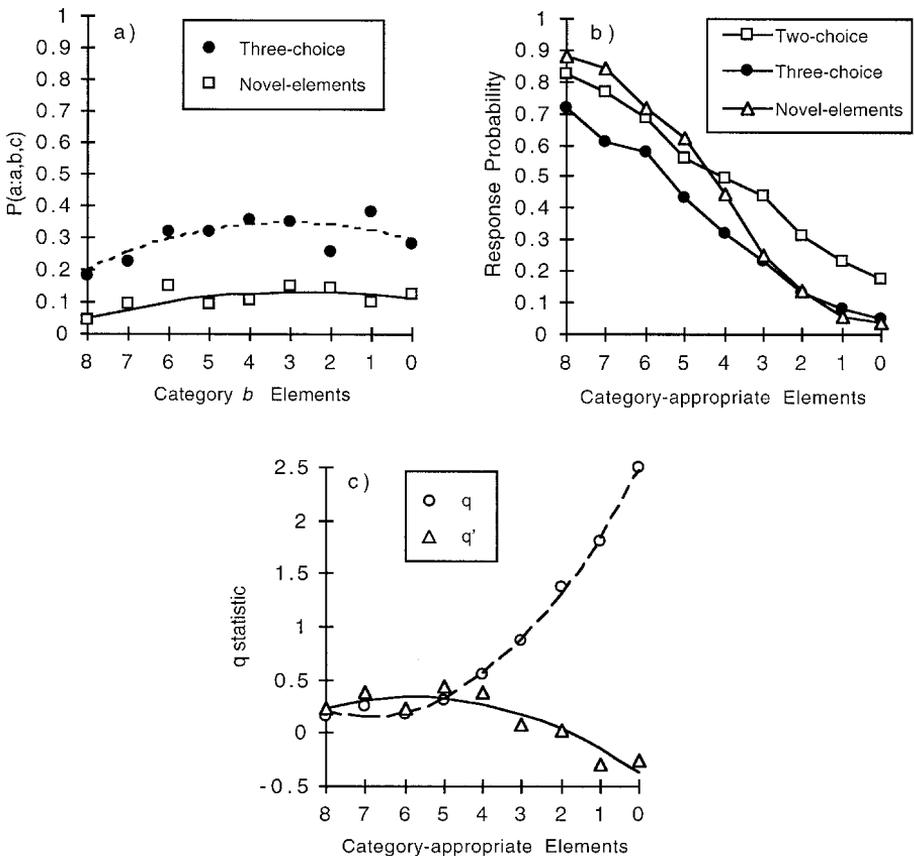
The subjects were allocated to one of three conditions such that an equal number (12) participated in each. These conditions were the two-choice and three-choice conditions of Experiment 1, plus a new novel-elements condition. The novel-elements condition differed from the three-choice condition only in the stimuli presented in the test phase (see Stimuli section). In all other respects, the procedure employed in Experiment 2 was identical to the procedure of Experiment 1, with the exception that, in Experiment 2, allocation of category labels to categories was varied across subjects by splitting the subjects in each condition into three equal-sized sub-groups. The allocation of labels to categories in these sub-groups is shown in Table 1.

TABLE 1  
Allocation of category labels in the three  
sub-groups of Experiment 2

		<i>Category</i>		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>Sub-group</i>	1	A	B	C
	2	B	C	A
	3	C	A	B

Results

Figure 4a shows the probability with which subjects in the three-choice and novel-elements conditions responded with the Category a label to test stimuli as a function of the number of Category b elements that they contained. The choice of Category b rather than Category c elements on the abscissa was arbitrary. For Category b there are equal numbers of subjects who receive “A”, “B”, and “C” as its label. The same holds for Category c, and hence response bias is controlled for whichever is chosen. Number of Category c elements equals eight minus the number of Category b elements, and so choosing Category c elements as the abscissa would simply reverse both functions left to right.



**Figure 4.** Results of Experiment 2. (a) Probability of producing a Category A response in the three-choice and novel-elements conditions as a function of the number of Category b elements in the presented test stimulus. (b) Mean response probability as a function of number of category-appropriate elements in each of the three conditions of Experiment 2. The ordinate is the mean of the probability of producing a Category B response to stimuli containing a given number of Category B elements and the probability of producing a Category C response to stimuli containing that number of Category C elements. (c) the  $q$  and  $q'$  statistics (see text) as a function of number of category-appropriate elements. Figures 4a and 4c show data as unconnected plot symbols; the lines are the best-fitting quadratics to the plotted data. Figure 4b shows data as connected plot symbols; there are no best-fit lines shown.

Inspection of Figure 4a suggests that both the  $P(a: a,b,c)$  function from the three-choice condition and the  $P(a: a,b,c)'$  function from the novel-elements condition show an inverted-U-shaped trend. To investigate this, second-order polynomial regression was performed on the nine mean data points of each function. The data from the three-choice condition was found to have the best-fit line  $P(a: a,b,c) = -0.006b^2 + 0.037b + 0.291$  (the dotted line in Figure 4a) where  $b$  is the number of Category  $b$  elements. This function was a significantly good fit to the data,  $F(2, 6) = 5.6, p < .05$ . The  $b^2$  coefficient was significantly different from zero,  $t(7) = 2.4, p < .05$ , as was the constant,  $t(7) = 8.0, p < .0005$ . The  $b$  coefficient was not significantly different from zero,  $t(7) = 1.7, p > .1$ . These analyses confirm that the  $P(a: a,b,c)$  function has an inverted-U trend. The best-fitting quadratic for the  $P(a: a,b,c)'$  function (from the novel-elements condition) was  $P(a: a,b,c)' = -0.003b^2 + 0.016b + 0.112$ , shown as a solid line in Figure 4a. This best-fit line also shows an inverted-U trend, but the function was not a significant fit to the data,  $F(2, 6) = 3.2, p > .1$ .

The data points plotted in Figure 4b are the average of the probability with which subjects responded with their Category  $b$  label to test stimuli with  $x$  Category  $b$  elements and the probability with which they responded with their Category  $c$  label to test stimuli with  $x$  Category  $c$  elements. In other words, like Figure 3a from Experiment 1, Figure 4b shows response probability as a function of number of category-appropriate elements. Averaging these two probabilities is appropriate because, across subjects, there is no factor that determines which of the two categories providing variable numbers of elements to test stimuli should be described as Category  $b$  and which as Category  $c$ .

For our current purposes it is not the data presented in Figure 4b that are of central interest, but the  $q$  and  $q'$  functions calculated from the mean data points that they display. These functions are shown in Figure 4c. The  $q$  function was calculated in exactly the same way as it was calculated in Experiment 1. In other words, for each value of category-appropriate elements, the mean three-choice probability was subtracted from the mean two-choice probability and the resulting number divided by the three-choice probability. For  $q'$ , the calculation was to subtract the three-choice probability from the novel-elements probability and to divide the resulting number by the three-choice probability.

Inspection of Figure 4c suggests that the  $q$  statistic shows an increasing trend whereas the  $q'$  statistic shows a decreasing trend. In order to specify these two functions more clearly, each set of nine data points were used to perform a second-order polynomial regression. The best-fit line for the  $q$  function was  $q = 0.049c^2 - 0.674c + 2.48$ , whereas for the  $q'$  function it was  $q' = -0.021c^2 + 0.244c - 0.368$ , where  $c$  is the number of category-appropriate elements. The best-fit line for the  $q$  function is shown as a dotted line in Figure 4c, whereas the best-fit line for the  $q'$  function is shown as a solid line. Both functions were a significant fit to the data,  $F(2, 6) = 803, p < .0005$ , for the  $q$  function and  $F(2, 6) = 17, p < .005$  for the  $q'$  function. The  $c^2$  coefficients for both functions were significantly different from zero,  $t(7) = 14, p < .005$ , for the  $q$  function and  $t(7) = 3.0, p < .05$ , for the  $q'$  function. The  $c$  coefficients were also significantly different from zero for both functions,  $t(7) = 24, p < .0005$ , for the  $q$  function and  $t(7) = 4.3, p < .004$ , for the  $q'$  function. Finally, both constants were significantly different from zero,  $t(7) = 2.5, p < .0005$ , for the  $q$  function, and  $t(7) = 3.8, p < .01$ , for the  $q'$  function.

## Discussion

The results of Experiment 2 directly contradict the predictions of the ratio rule acting on the output of a magnitude-based model of categorization, within the assumption that magnitude is a univariate function of category-appropriate elements.

Under this assumption, our results pose two central problems for the ratio rule. First, the ratio rule predicts that the  $q$  and  $q'$  statistics should show the same direction of change over any interval of category-appropriate elements. However, the best-fitting quadratics for these functions show opposite directions of change. Second, the ratio rule predicts that the  $q$  function and  $P(a: a,b,c)$  function for the three-choice condition should also show the same direction of change over any interval of category-appropriate elements. Irrespective of whether one chooses to plot  $P(a: a,b,c)$  as a function of Category b elements or as a function of Category c elements, the fact that the best-fitting quadratics for the  $q$  and  $P(a: a,b,c)$  functions are of opposite shape (U vs. inverted-U shape) is contrary to the predictions of the ratio rule. One argument against this conclusion might be that  $P(a: a,b,c)$  is an increasing, decelerating function rather than an inverted-U function. As  $q$  is an increasing, accelerating function, both functions show the same direction of change, and so the predictions of the ratio rule are upheld for these data. However, closer study of Equations 3 and 6 reveals that both functions must be accelerating or decelerating over any given interval. Variability in both functions is determined by changes in the term  $(v_B + v_C)$ . If the rate at which this term changes is increasing over a given interval then both  $q$  and  $P(a: a,b,c)$  must show an increasing rate of change over that interval.

Experiment 2 also substantially replicates the results of Experiment 1. In both experiments the best-fitting quadratic for the  $P(a: a,b,c)$  function shows an inverted-U trend, whereas the best-fitting quadratic for the  $q$  function shows a U-shaped trend. Any cubic trend in the  $q$  statistic is not replicated in Experiment 2.

## MODELLING

Wills and McLaren (1997) proposed a winner-take-all (WTA) connectionist model as an alternative to the ratio rule in magnitude-based models of categorization. The work presented in this section demonstrates that our WTA model, unlike the ratio rule, can account for the results presented in this paper under the assumption of univariate magnitude functions.

The WTA system is illustrated in Figure 5. The magnitude terms for each category are passed to separate units in the network as input activations. The input activations are constrained to be no greater than one and no less than zero. The output activity of each unit in the WTA system is a function of the total input that it receives. In addition to the magnitude-term inputs, each unit in the WTA system has a fixed excitatory connection to itself and fixed inhibitory connections to other units. These connections cause the units to "compete" with one another until only one has a non-zero activation. In our system, a decision is assumed to be reached when the activation of one unit exceeds that of its nearest competitor by some threshold value. The category whose unit has the highest activity is chosen at that point. Network systems based on the concepts of mutual inhibition and self-excitation have been proposed previously by, amongst others, Grossberg

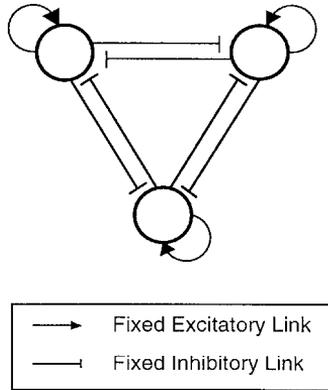


Figure 5. The winner-take-all model.

(1976), McClelland (1979), Page (in press), and Rumelhart and Zipser (1986). They have also been previously employed as models of specific experimental paradigms; such as serial recall (Houghton, 1990) and simple binary-choice reaction time studies (Usher & McClelland, 1995).

The input activations to our WTA system are assumed to be noisy. In other words, although the magnitude term represents the mean level of input activity, the momentary level of input varies randomly about this mean. This means that, although the category with the highest magnitude term is most likely to be chosen, all other categories have a finite chance of being chosen. Further, it means that the WTA model can be considered as a connectionist implementation of the basic principles of Thurstonian choice (see Introduction).

### Detailed specification of the WTA system

Prior to the presentation of a stimulus, the output activation of all units is assumed to start from zero. Once the magnitude terms have been presented to the WTA system, the output activation of each unit is updated repeatedly until a decision has been reached. The output activation of unit  $i$  in the WTA system on update  $c$  is determined by

$$o_{i,c} = \frac{o_{i,c-1} + En_{i,c}}{1 + En_{i,c} + D} \tag{10}$$

if  $n_{i,c} > 0$  and otherwise by

$$o_{i,c} = \frac{o_{i,c-1} + En_{i,c}}{1 - En_{i,c} + D} \tag{11}$$

where  $n_{i,c}$  is the total input to unit  $i$  on update  $c$ , and  $E$  and  $D$  are parameters representing the rate of excitation and decay of activation within the unit. Equations 8 and 9 can be described as *time-averaging*, *squashing functions*. They are squashing functions because whatever the absolute magnitude of the input to a unit, its activity is constrained by these functions to fall between +1 and -1. The equations are time averaging because the

inclusion of the unit's activity on the previous update in the calculation of its current activation means activation represents, not the instantaneous input to a unit, but an average of the inputs it has received in the recent past. A more detailed discussion of the behaviour of activation functions of this general form is provided by Grossberg (1976). In our simulations  $o$  is constrained to be non-negative.

The total input to a unit  $i$  on update  $c$  is given by

$$n_{i,c} = r_{i,c} + o_{i,c-1} - \sum_{j \neq i} o_{j,c-1} \quad 12$$

The term  $r_{i,c}$  in this equation is the value of the noisy input produced by the magnitude term  $v_i$  and presented to unit  $i$  on update  $c$ . In the simulations that follow the noise added to  $v_i$  ranges from  $+N$  to  $-N$ , has a mean of zero, and has a rectangular distribution (i.e., all values from  $+N$  to  $-N$  are equally likely). Superimposed on this noise function is the constraint that  $r_{i,c}$  cannot exceed one or fall below zero. The remaining terms of Equation 12 state that each unit receives a positive input equal to its own activation on the last update and a negative input equal to the sum of the activations of the other units. It is these terms that specify the self-excitatory and mutually inhibitory properties of the WTA system.

The final component of the WTA system is the *decision threshold*. The decision threshold is the amount by which the unit with the highest activation has to exceed the activation of its nearest competitor in order to "win" and cause the production of its associated response. This value is the  $S$  parameter of the WTA model. In the current simulation,  $S$  is set to 0.18 for the two-choice condition, 0.65 for the three-choice condition, and 0.72 for the novel-elements condition. Employing a different value of  $S$  for each condition is in line with previous applications of the model where we have assumed that both the type of decision (two-choice vs. three-choice here), and the presence of novel elements in test stimuli, affects the value of  $S$  (Wills, 1998; Wills & McLaren, 1997). The use of a higher value for  $S$  in conditions where some elements of the test stimuli are novel represents a hypothesis that subjects trade speed for accuracy when asked to make decisions about stimuli that are unfamiliar.

The remaining parameters,  $E$ ,  $D$  and  $N$ , are set to 0.2, 0.1, and 1.1, respectively. These values are the same as those employed by Wills and McLaren (1997) in the simulation of their experiments, and by Wills (1998) in the simulation of the experiments presented in Jones et al. (1998).

## Specification of magnitude terms

In order to simulate the results of Experiment 2, magnitude terms for each category at each value of category-appropriate elements are required. We will assume for the purposes of this simulation that these magnitude terms are linear functions of number of category-appropriate elements. Each category is assumed to have the same magnitude function, which, in the current simulations, takes the form

$$v_i = 0.047c_i + 0.012 \quad 13$$

where  $c_i$  is the number of category  $i$  elements that the stimulus contains.

In previous applications of the WTA model, we have employed a single-layer, elemental, error-correcting network to specify the magnitude terms. This model was employed because it captured something of the effects of varying the number of training examples presented. In the current experiments, because number of training examples does not vary between conditions, the performance of this network model is characterized by a single linear function. Equation 13 specifies this function for the network model presented in Wills and McLaren (1997) when the model's learning-rate parameter is set to 0.0025. This value is of the same order of magnitude as learning rates that we have employed previously in simulating experiments of this sort.

## Simulation

The simulation we present is of Experiment 2. Experiment 1 is not simulated because its two conditions are repeated in Experiment 2 and because possible response bias is not controlled for in Experiment 1.

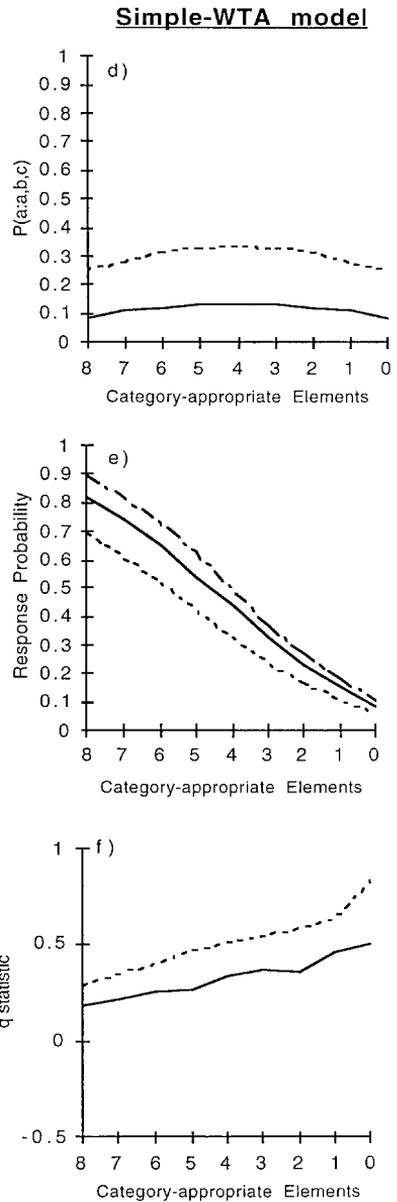
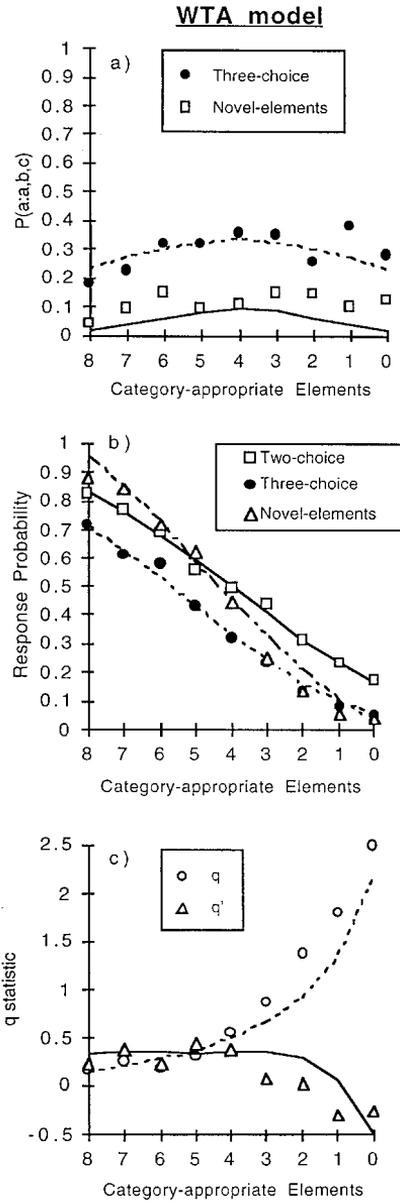
The simulation involved presenting, in turn, every combination of magnitude terms that would be produced by the application of Equation 13 to the set of test stimuli employed in the experiment. Each combination was presented 50,000 times, and the decision made by the WTA system was recorded each time. The probability with which each response was made to each type of test stimulus was then calculated from this record. This was done separately for each of the three experimental conditions.

In the two-choice condition of our experiment, subjects were not allowed to make Category a responses. In our WTA model this was simulated by fixing the output activation of the Category a unit ( $o_a$ ) at zero. The output activations of the Category b and c units were allowed to take the values determined by Equations 10, 11, and 12. The assumption made in doing this is that only allowed responses compete for the right to produce a response. In the three-choice condition and in the novel-elements condition, all three output activations were allowed to vary. In the novel-elements condition,  $v_A$  was assumed to be zero.<sup>2</sup> However, the noise component of the magnitude term means that  $o_a$  will not always be equal to zero, and hence Category a will be chosen with a non-zero probability.

The response probabilities predicted by the WTA model are presented (as lines) in Figures 6a–6c, along with the mean probabilities observed in Experiment 2 (as unconnected plot symbols). It can be seen from this figure that the WTA model, unlike the ratio rule, can predict the correct shape for the  $q$ ,  $q'$ ,  $P(a: a,b,c)$  and  $P(a: a,b,c)'$  functions. Further, the data and predictions correspond fairly closely, although it may be noted that the values predicted for  $q$  and  $P(a: a,b,c)'$  are slightly lower than those observed. Nevertheless, our simulation demonstrates that the WTA model is capable of predicting the major trends observed in our experiment. We have already determined that (within certain assumptions) the ratio rule is unable to do so.

---

<sup>2</sup> Inspection of Equation 13 might suggest that  $v_a$  should be set to 0.012 rather than zero, on the grounds that  $c_a$  will be equal to zero. However, for the simple network model presented in Wills and McLaren (1997),  $v_a$  will be equal to zero because the novel elements are not presented in the training phase, and hence no association will form between them and the category labels. The predictions presented are not dependent on the choice of zero rather than 0.012 for the value of  $v_a$ .



**Figure 6.** Simulation of Experiment 2 with the full WTA model (Figures 6a–6c) and the simple WTA model (Figures 6d–6f). The lines show the predictions of the models. The observed data of Experiment 2 are re-plotted as unconnected plot symbols in Figures 2a–2c for comparison.

## Discussion and further simulation

The WTA model is a relatively complex system with four free parameters ( $E$ ,  $D$ ,  $N$ , and  $S$ ). The ratio rule, in contrast, has no free parameters—its predictions are entirely determined by the magnitude terms that it is presented with. This contrast in complexity raises two related questions. First, is it simply increased complexity and, in particular, the presence of more parameters, that permits the WTA model to successfully account for our data? Second, which components of this complex model are central to producing its predictions, and which are unnecessary for predicting the current data set?

Despite its complex statement, a simple theoretical principle lies at the heart of the WTA model. The principle is that one always chooses the allowed alternative with the largest magnitude term. If the alternative with the largest magnitude term is not chosen, this is because magnitude terms are noisy and the alternative chosen appeared to have the largest magnitude term over the finite time allowed for decision.

The most basic instantiation of the central “pick the biggest” principle underlying the WTA model would be to take the instantaneous value of each of the noisy magnitude terms and pick the alternative with the largest value at that instant. The probability with which this *simple-WTA* system picks each alternative is entirely determined by the means and distributions of the magnitude terms. Like the ratio rule, the simple-WTA system has no free parameters, although more information about the magnitude terms is required to derive predictions from the simple-WTA system.

We have calculated the predictions of the simple-WTA system assuming magnitude terms with a rectangular distribution with a width of 0.7 and means determined by Equation 13. The correspondence between predictions and data is by no means as good for this simplified model as it is for the full model (see Figures 6d–6f). However, it can be seen that the simple model can correctly predict the basic trends observed in the  $q$ ,  $P(a: a, b, c)$  and  $P(a: a, b, c)'$  functions. The  $q$  function is predicted to show an increasing trend, whereas the  $P(a: a, b, c)$  and  $P(a: a, b, c)'$  functions are predicted to show an inverted-U-shaped trend. Employing Gaussian distributions with a standard deviation of 0.28 (which have a rectangular equivalent with a width of 0.7) produces comparable results.

The fact that the simple-WTA model can predict the  $q$  function to be different in shape from the  $P(a: a, b, c)$  and  $P(a: a, b, c)'$  functions is encouraging. It means that the success of the WTA class of models in predicting these functions where the ratio rule failed to do so is not simply due to the greater complexity of WTA models or the greater number of parameters employed. It also suggests that it is the central “pick the biggest from noisy alternatives” principle, which underlies the success of the full WTA model, rather than the specific way this principle is instantiated.

However, the presented simulation of the simple-WTA model does not correctly predict the trend in the  $q'$  statistic. We cannot state definitively that the simple-WTA model is unable to predict this trend, but we believe that it is the adoption of different decision thresholds, for the three-choice and novel-elements conditions in the full model that allows it to correctly predict the  $q'$  data. Were  $S$  to take the same value in both conditions, the probability of choosing, say, the Category b response for a stimulus with  $x$  Category b elements would always be greater in the novel-elements condition than in the three-choice condition (and hence  $q'$  would always be positive). This is because there is

less competition from the Category a alternative in the novel-elements condition (due to the stimuli in that condition containing no Category a elements). Increases in the value of  $S$  reduce the chance of choosing an alternative that does not have the highest magnitude term. Hence increasing the value of  $S$  in the novel-elements condition has the effect of reducing the response probabilities in the region three category-appropriate elements down to zero category-appropriate elements. At zero category-appropriate elements, the effect is sufficiently great to make the predicted probability of response in the novel-elements condition lower than the predicted probability for the three-choice condition, causing  $q'$  to be negative and producing a downward trend in the  $q'$  statistic.

## GENERAL DISCUSSION

Any theory of learning and memory whose output is a set of magnitude terms must specify how these terms translate into testable predictions. Where those predictions concern response probabilities, it is commonly assumed that the ratio rule provides the appropriate translation. With certain qualifications, this assumption has been shown to be incorrect for the categorization experiments presented in this paper.

If any one step of a chain of inferences is incorrect then the conclusions drawn from that process must be brought into question. Consequently, theoretical conclusions about the nature of learning and memory that have been drawn from models that employ the ratio rule must be re-examined if our conclusion is valid. Conversely, if the assumptions we have made in coming to our conclusions can be shown to be invalid then the ratio rule is not necessarily incorrect. We consider some possible arguments against our conclusions.

One general class of argument arises from the fact that we have estimated the shape of our four functions,  $q$ ,  $q'$ ,  $P(a: a,b,c)$ ,  $P(a: a,b,c)'$ , from mean data rather than from the data of individual subjects. On this basis, one could argue that although we have demonstrated the ratio rule to be incorrect for average responses, it may actually be correct for individuals. If this were shown to be true it would not substantially change our conclusion that the ratio rule is incorrect, as most models of memory have been predominantly applied to mean rather than individual data.

Another variant of the general argument against using mean data would be that the assessment of functions provided is not reliable or is substantially inferior to that provided by multivariate methods. Although the  $q$  statistic appears to have a cubic component in Experiment 1 but not in Experiment 2, the presence of the same basic trends in these two experiments provides substantial evidence against the position that our assessment method is unreliable. Further, in a between-subjects design, the adoption of multivariate methods for the  $q$  statistic would be inappropriate because the function is calculated from decisions made by different people. This problem could be addressed by the adoption of a within-subjects design. The central problem with this approach would be that if any subject made exclusively Category a responses at any value of category-appropriate elements in the three-choice decision, then the value of  $q$  calculated from their data would be undefined, because  $P(b: a,b,c)$  would be estimated as zero, and any number divided by zero is undefined. There seems to be no satisfactory way of addressing this problem if it occurred—excluding subjects because one's dependent measure cannot represent their

behaviour is not really appropriate, and replacing the undefined values with some real number would be difficult to do in a manner that was both non-arbitrary and relatively atheoretical.

A reasonable criticism of our experiments would be to point out that the stimuli employed are rather more complex than those typically used in category-learning experiments. It may therefore be argued that the results we have found with our complex stimuli do not generalize to more simple stimuli, such as simple outline drawings or basic geometric forms. If this were shown to be true, the use of the ratio rule to describe categorical decisions about simple stimuli might be valid. If one accepts that the objects we have to categorize are often complex, a theory of categorical decision that can only explain our decisions about stimuli that are simple seems to be of limited use. However, the question of whether our results generalize to simple stimuli is clearly an empirical one and worthy of investigation.

A different line of argument against our conclusion that the ratio rule is incorrect would be to demonstrate that, for specific models of categorization, the resultant magnitude terms for our test stimuli were not univariate functions of the number of category-appropriate elements. For example, magnitude terms in the generalized context model (GCM: Nosofsky, 1986) are determined by the position of stimuli in a psychological similarity space. If one could determine the position of our test stimuli in such a space and demonstrate that they were not at a constant distance to the Category *a* prototype then the GCM's use of the ratio rule would potentially be appropriate. This conclusion is itself subject to two qualifications. First, it would have to be demonstrated that the GCM could actually predict our results from the psychological similarity space derived. Second, the positions of stimuli in psychological similarity space are determined in GCM by a model of identification that assumes the ratio rule to be correct. Some way of avoiding the circularity involved in trying to test a theory whilst assuming it to be true would have to be found.

A more general argument against our conclusions would be to state that magnitude terms are affected not only by the stimulus presented but also by the alternatives allowed for the decision. If this were true for our experiments then the derivation of the *q* statistic presented in Equations 4, 5, and 6 would be flawed, as  $v_B$  is specifically assumed not to be affected in this way. This argument may be seen as a qualification of our conclusions, which may thus be stated more fully as "the ratio rule is incorrect for models that have no process by which information about allowed decision alternatives can affect the magnitude terms produced". This qualification excludes none of the categorization models cited in this paper from our conclusions.

Interestingly, a number of theorists have proposed ways in which the ratio rule as originally stated may be developed to allow it to be sensitive to the alternatives allowed for decision. For example, Restle (1961) suggests that, in a two-alternative decision, both magnitude terms should have a value subtracted from them that represents the similarity of those two alternatives. Hence, under Restle's theory, the decision between a foreign holiday and a foreign holiday plus a peanut, is treated as a decision between a peanut and nothing because the holiday is received irrespective of one's choice. Restle did not extend his theory to situations involving more than two alternatives and hence it cannot be applied to the current data. Tversky (1972) proposed a similar theory, which is applicable

to the  $N$ -alternative case, but his theory replaces the single magnitude term per alternative required by the ratio rule with a set of magnitude terms that represent the importance of particular aspects of each alternative. Most models that employ the ratio rule would require substantial revision in order to allow them to provide the set of magnitude terms required.

If one needs to replace the ratio rule in its basic form with a much more complex theory of decision making, such as Tversky's, in order to explain our data, then this would support our central conclusion that the ratio rule, as currently employed, is incorrect. However, an alternative approach might be to make some minor modification to the ratio rule that would allow it to account for our results. Although it is impractical to consider every possible modification, we discuss one particular idea which has been proposed by a number of researchers (e.g., Aitken, 1996; Nosofsky & Zaki, 1998).

The modification is to add a constant, which can be thought of as representing the presence of background noise in the decision process. Specifically, the ratio rule is modified to

$$P(i) = \frac{v_i + X}{\sum_{j=1}^n v_j + nX} \quad 14$$

where  $X$  is a constant representing background noise. This modification allows the ratio rule to account for (amongst other things) the results of Jones et al. (1998) even if linear magnitude functions are assumed. However, the addition of a constant will not affect the prediction that change in  $P(a: a,b,c)$  and change in  $P(a: a,b,c)'$  are determined by change in  $(v_B + v_C)$ . Similarly, this modified ratio rule results in a prediction for  $q$  of

$$q = \frac{v_A + X}{v_B + v_C + 2X} \quad 15$$

and for  $q'$  of

$$q' = \frac{v_A - v_N}{v_B + v_C + v_N + 3X} \quad 16$$

The addition of constants will not affect the prediction that changes in  $q$  and  $q'$  are determined by the changes in  $v_B + v_C$ . Hence, this modification still predicts that all four empirical functions,  $q$ ,  $q'$ ,  $P(a: a,b,c)$  and  $P(a: a,b,c)'$ , will have the same direction of change over any given interval of category-appropriate elements, which we have shown is not the case.

Our central conclusion is that the ratio rule is an inappropriate theory of categorical decision and should be replaced by a system based on the principles of Thurstonian choice. However, the ratio rule and Thurstonian choice need not necessarily be considered as different classes of explanation. As discussed earlier, Yellott (1977) demonstrated that the predictions of the ratio rule are equivalent to a Case V Thurstonian choice process with double exponential noise distributions. As such, the ratio rule may be considered as a description of one member of the set of Thurstonian choice

processes.<sup>3</sup> If the ratio rule is considered in this way then our central conclusion is more properly stated as “the Case V double exponential Thurstonian choice process is an inappropriate model of categorical decision, but other Thurstonian choice processes are potentially appropriate”. However, one might alternatively consider the ratio rule to be a statement that people make probabilistic judgements on the basis of deterministic magnitude terms, in contrast to the Thurstonian theory that people make deterministic judgements on the basis of probabilistic magnitude terms. If considered in this manner, the ratio rule and Thurstonian choice are clearly different classes of explanation.

Finally, it is important to state that the WTA model presented by Wills and McLaren (1997) is just one of a large class of models that produce predictions about response probability and response time on the basis of a competitive race. Other examples of this class include a number of connectionist models (e.g., Houghton, 1990; Lacouture & Marley, 1991; Usher & McClelland, 1995), some instance-based memory models (e.g., Logan, 1988; Nosofsky & Palmeri, 1997), and several other mathematical models of various descriptions (e.g., Karpiuk, Lacouture, & Marley, 1997; Ratcliff, 1978). Page (in press) provides an excellent discussion of the similarities and differences between some of the aforementioned theories. By using our WTA model to simulate the data presented in this paper we do not intend to imply that it is the only model of its class that has the potential to explain our results. Indeed, the partial success of our simple WTA model suggests that it is the general principles of Thurstonian choice, rather than the competitive race itself, that underly the success of our full model. Many models employing these general principles are likely to be able to explain many of our results (e.g., Ashby & Townsend, 1986), as long as the noise distribution employed does not render their predictions indistinguishable from those of the ratio rule.

## REFERENCES

- Aitken, M.R.F. (1996). *Peak shift in pigeon and human categorization*. Unpublished PhD thesis, University of Cambridge, Cambridge, U.K.
- Anderson, J.R. (1991). A rational analysis of categorization. *Psychological Review*, 98, 409–429.
- Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.
- Bradley, R.A. (1954). Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparison. *Biometrics*, 10, 375–390.
- Burke, C.J., & Zinnes, J.L. (1965). A paired comparison of pair comparisons. *Journal of Mathematical Psychology*, 2, 53–76.
- Clarke, F.R. (1957). Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 29(6), 715–720.
- Estes, W.K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
- Gluck, M.A., & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117(3), 227–247.
- Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.

---

<sup>3</sup> Considering the ratio rule in this way assumes that Gaussian distributions are not a defining property of Thurstone's theory. Such an assumption seems reasonable given that Thurstone argued that “the only valid justification for bringing in the probability curve [i.e., the Gaussian distribution] . . . is that its presence can be experimentally tested” (Thurstone, 1927, p. 373).

- Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121–134.
- Hohle, R.H. (1966). An empirical evaluation and comparison of two models for discriminability scales. *Journal of Mathematical Psychology*, 3, 174–183.
- Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 418–439.
- Hopkins, J.W. (1954). Incomplete block rank analysis: Some taste test results. *Biometrics*, 10, 391–399.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation*. London: Academic Press.
- Hull, C.L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, No. 123.
- Jones, F.W., Wills, A.J., & McLaren, I.P.L. (1998). Perceptual categorization: Connectionist modelling and decision rules. *Quarterly Journal of Experimental Psychology*, 51B(3), 33–58.
- Karpiuk, P., Jr., Lacouture, Y., & Marley, A.A.J. (1997). A limited capacity, wave equality, random walk model of absolute identification. In A.A.J. Marley (Ed.), *Choice, decision and measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kruschke, J.K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(1), 3–26.
- Lacouture, Y., & Marley, A.A.J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science*, 3(4), 401–433.
- Lamberts, K. (in press). Information-accumulation theory of speeded categorization. *Psychological Review*.
- Logan, G.D. (1988). Towards an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Luce, R.D. (1959). *Individual choice behavior*. New York: John Wiley & Sons.
- McClelland, J.L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.
- McClelland, J.L., & Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114(2), 159–188.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207–238.
- Nosofsky, R.M. (1986). Attention, similarity and the identification-categorisation relationship. *Journal of Experimental Psychology: General*, 115(1), 39–57.
- Nosofsky, R.M., & Palmeri, T.J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2), 266–300.
- Nosofsky, R.M., & Zaki, S.R. (1998). Dissociations between categorization and recognition in amnesiacs and normals: An exemplar-based interpretation. *Psychological Science*, 9(4), 247–255.
- Page, M. (in press). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*.
- Pearce, J.M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- Pollack, I., & Decker, L. (1960). Consonant confusions and the Constant Ratio Rule. *Language & Speech*, 3, 1–6.
- Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3), 353–363.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 97, 285–308.
- Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Current research* (pp. 64–99). New York: Appleton-Century-Crofts.
- Restle, F. (1961). *Psychology of judgement and choice*. New York: Wiley.

- Rumelhart, D.E., & Zipser, D. (1986). Feature discovery by competitive learning. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.
- Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.
- Thurstone, L.L. (1927). A law of comparative judgement. *Psychological Review*, *34*, 273–286.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299.
- Usher, M., & McClelland, J.L. (1995). *On the time course of perceptual choice: A model based on principles of neural computation* (Tech. Rep. PDP.CNS.95.5). Carnegie Mellon University.
- Wagner, A.R. (1981). SOP: A model of automatic memory processing in animal behavior. In N.E. Spear & R.R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Wills, A.J. (1998). *Categorization: From features to decisions*. Unpublished PhD thesis, Cambridge University, Cambridge, UK.
- Wills, A.J., & McLaren, I.P.L. (1997). Generalization in human category learning: A connectionist explanation of differences in gradient after discriminative and non-discriminative training. *Quarterly Journal of Experimental Psychology*, *50A*(3), 607–630.
- Yellott, J.I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–144.

*Original manuscript received 31 July 1998*  
*Accepted revision received 11 November 1999*