Adapting to a Response Deadline in Categorization

A. J. Wills (a.j.wills@ex.ac.uk)
School of Psychology, University of Exeter,
Perry Road, Exeter. EX4 4QG. England

Abstract

The effect of a response deadline on categorical decisions was investigated. Time available for response was manipulated in the test phase, along with stimulus difficulty. Effects of these manipulations were observed in response accuracy, and in the mean, standard deviation and skew of the reaction times. The effects observed demonstrate that participants responded to the deadline in an adaptive manner - reducing their reaction time to long-latency decisions whilst leaving short latency decisions relatively unaffected. A simple connectionist model of categorical decisions (Wills & McLaren, 1997) is shown to account for this behavior.

Introduction

Categorization is a basic and essential cognitive function. Our ability to engage it has been well studied, and a number of different theories of the underlying processes have been proposed (e.g. Ashby & Gott, 1988; Gluck, 1991; Nosofsky, 1986; Nosofsky, Palmeri & McKinley, 1994). At first, attempts to quantitatively fit models of categorization to empirical data concentrated on categorization accuracy. However, in recent years, models which have the potential to predict reaction time distributions in categorization have been developed and evaluated (e.g. Ashby, 2000; Maddox, Ashby & Gottlob, 1998; Lamberts, 2000; Nosofsky & Palmeri, 1997; Wills & McLaren, 1997).

This paper focuses on the effects of imposing a response deadline on a) participants' response accuracy and b) the nature of their reaction time distributions. It has been known for some time that categorical decisions made under time pressure may be different to those made without time pressure (eg. Smith & Kemler Nelson, 1984). More recently, this avenue of research has been developed by investigation of the effects of time pressure with more complex stimuli (e.g. Lamberts, 1995; Palmeri & Blalock, 2000) coupled with formal modeling of the results found (e.g. Lamberts, 1995).

It is worth considering Lambert's (1995) study in a little more detail as it provides one motivation for the current work. At one level, the results found are intuitive. In these experiments, Lamberts employed a simple deadline procedure. Participants first learned, in the absence of time pressure, to categorize artificial stimuli (schematic faces) into two categories. Following this training, participants had to categorize test stimuli before a given deadline (e.g. 1600ms from stimulus onset). Failure to respond in time resulted in an error tone, followed by the presentation of the next stimulus. Participants were informed about the time available for response, which changed at regular intervals. In one experiment, the deadlines employed were 600ms, 1100ms, 1600ms and no deadline. Participants were less accurate at shorter deadlines. Interestingly, the effect was stimulus specific, with some stimuli being considerably more affected by time pressure than others. Lamberts proposed a particular formal model of this effect (the "Extended Generalized Context Model" or EGCM, Lamberts, 1995) and showed that it provided a good fit to the accuracy data.

Time pressure and reaction time

Lamberts' experiments reveal another result. In his experiments, categorization in the absence of a response deadline takes approximately 1500ms (Lamberts, 1995, experiment 2). As the stringency of the deadline increases, so the mean reaction times decrease, with categorization under a 600ms deadline taking about 450ms. In other words, categorical decisions appear to take considerably less time when there is time pressure than when there is not. This is, of course, intuitively obvious. The interest, from the perspective of the current paper, is that there seem to be at least three distinct reasons why it might happen. When considering the following, it is important to remember that the descriptions relate to observed reaction time distributions - they are not statements about underlying process:

Non-selective adaptation: The participant reacts to the imposition of the deadline in a manner that decreases all reaction times in the distribution by a fixed amount. As a consequence, mean of the distribution will drop, but the standard deviation and skew will be unaffected.

Selective, linear adaptation: The participant reacts to the imposition of the deadline in a manner which decreases all reaction times in the distribution by a fixed factor (i.e. $RT_{deadline} = f \times RT_{no\ deadline}$). As a result, the mean and standard deviation of the distribution will drop, but the skew will be unaffected.

Selective, non-linear adaptation: The participant reacts to the imposition of the deadline in a way that cannot be characterized as non-selective, or selective, non-linear, by the definitions above. Changes in the mean, standard deviation, and skew of the distribution may all be observed.

Demonstrating adaptation to a deadline

It therefore seems clear that to distinguish between these explanations, one must estimate changes in the mean, standard deviation and skew of the reaction time distribution produced by imposition of a deadline. Whilst the experiment reported in this paper is by no means the first to investigate the effects of a deadline on categorization accuracy and reaction time, previous work has had at least one of the two following limitations:

Missing data artifact

In a number of studies (e.g. Lamberts, 1995; Lamberts & Brockdorff, 1997; Palmeri & Blalock, 2000) it is possible that the changes observed are an artifact of the data collection procedure. In a response deadline procedure, longer-than-deadline responses typically result in a "time out" error and hence no data about reaction time is available for that trial. As a direct consequence, mean response time is lower than it would have been without a deadline. The same problem applies to studies that compare two different deadlines. In experiments where percentage of time-outs is reported by condition, they can been seen to increase as the response deadline becomes more stringent.

One solution to this problem is to use a "response signal" procedure (e.g. Lamberts, 1998) where participants are instructed to respond as soon as possible after they get a signal to do so. Another solution (see e.g. van Zandt, Colonius & Proctor, 2000) is to provide a "too slow" signal after the response has been made.

A third possibility is to use the standard response deadline procedure, but only evaluate responses that fall below a certain percentile of the reaction time distribution (with time-outs being considered as the slowest trials). The largest number of time-outs made at any level of time pressure, by any participant, to any of the test stimuli, determines this percentile. For all conditions and stimuli, only responses that fall below that fixed percentile are considered. It is therefore important to keep the percentage of time-outs low so a reasonable amount of data is still available for analysis. It is this final possibility that is employed in the current study.

Insufficient information

The three possibilities for adaptation outlined above can only be distinguished if one has estimates for the mean, standard deviation, and skew of the reaction time distributions. Recently, many studies of categorization have begun to report reaction time distributions in detail (e.g. Maddox & Ashby, 1996). However, categorization studies that employ time pressure as a manipulation tend to concentrate on categorization accuracy, and may also report mean reaction times. Data from different tasks, such as perceptual matching, show that the mean, standard deviation, and positive skew all reduce in response to increasing time pressure (van Zandt et al., 2000).

Given the absence of appropriate information, it was decided to perform a short empirical study that would have the potential to discriminate between the three types of adaptation to a response deadline which have been outlined. This is followed by a demonstration that a particular model of categorical decisions (Wills & McLaren, 1997) can mimic the results found. Implications of both the empirical and the theoretical investigations for categorization research are then discussed.

Experiment

The current experiment had two phases. In the training phase, participants were presented with novel, abstract stimuli paired with either the category label "A" or the category label "B". In the test phase that followed, participants had to decide the category membership (A or B) of unlabelled stimuli either without time pressure, with a 2500ms time limit for each decision or with a 1000ms time limit for each decision (a between-participants manipulation).

Whilst these deadlines may appear relatively lax compared to the reaction times observed in some classification tasks, previous work (e.g. Wills & McLaren, 1997) indicates they represent a fairly high level of time pressure for participants with relatively little experience of the complex stimuli employed.

Method

Participants and apparatus

The participants were 44 adults, mainly undergraduate students. The experiment was in two different, quiet cubicles on two Acorn RISC PC computers, with 14" color monitors. Participants sat 1 meter from the screen.

Stimuli

Each stimulus was a collection of twelve different small pictures (hereafter "elements") in a 4.5cm by 3.5cm rectangle outline, arranged on an invisible four-by-three grid (see Figure 1 for an example stimulus). Every stimulus contained twelve elements drawn from the pool of thirty-six elements we have used in previous experiments (see Jones, Wills & McLaren, 1998, p.37). At the beginning of the experiment, and separately for each participant, 12 elements from the pool were randomly designated as category A elements, and a different 12 as category B elements. The remaining 12 elements were not used for that participant.

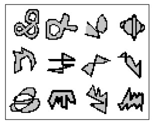

Figure 1: An example stimulus

Training stimuli

Sixty training stimuli (thirty from each category) were created for each participant. Each training stimulus was constructed by starting with all 12 elements characteristic of a particular category (e.g. category A elements for a category A training stimulus). Then, each element in the training stimulus underwent a 10% chance of being replaced by an element chosen from the other set (e.g. replaced by a category B element in the case of a category A training stimulus). Choice of replacement elements was random within the constraint that no element could occur more than once in any given stimulus. 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.

This method of stimulus construction produces training examples which are composed predominately of elements characteristic of a particular category but which also exhibit considerable variability.

Test stimuli

Test stimuli were designed to vary in difficulty of categorization. Given the nature of the training stimuli, the correct response to a test stimulus is to categorize it as an "A" if it contains more A elements than B elements, and as a "B" otherwise. A number of previous experiments have demonstrated that as the difference between the number of A elements and the number of B elements increases in a stimulus of this type, the probability of a correct classification also increases (Jones et al, 1998; Wills & McLaren, 1997). Test stimuli in this experiment are therefore described in terms of their difference scores (the absolute value of the number of A elements minus the number of B elements).

All stimuli contained twelve elements, so there are seven possible difference scores and hence seven levels of difficulty. The seven differences scores are 12, 10, 8, 6, 4, 2 and 0, which are denoted as having a difficulty level of 1,2,3,4,5,6 and 7 respectively. Twenty examples at each of the first six levels of difficulty were created for each participant. The specific elements used to create each test stimulus were chosen randomly within the constraint provided by the difference score, and the constraint that stimuli in which category A elements were more numerous that category B elements should occur with the same frequency as stimuli in which category B elements were more numerous than category A elements. As in the construction of the training stimuli, the position of elements within a test stimulus was randomly determined, and no element was allowed to occur more than once in any given stimulus.

Ten examples of stimuli with a zero difference (difficulty level 7) were also generated for each participant. However, as there is no correct answer for such stimuli, performance on them is not analyzed in this paper.

Procedure

Participants were allocated to one of three groups that differed only in the time allowed for decision in the test phase. These groups are referred to hereafter as the *1000ms, 2500ms* and *No-deadline* groups. Sixteen participants were allocated to the 1000ms group, sixteen to the 2500ms group, and twelve to the no-deadline group.

The sixty training stimuli were presented sequentially and in a random order. Each example was presented for five seconds in the center of the monitor accompanied by the appropriate category label (presented as a large capital A or B in an outline rectangle immediately to the right of the stimulus). The stimulus and the category label were then replaced with mid-gray rectangles that stayed on the screen for two seconds and were followed by the next example. Participants were not required to respond in any way in this first phase of the experiment but were 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 other experiments (Jones et al, 1998; Wills & McLaren, 1997).

The training phase was followed immediately by the test phase. There were 130 stimuli in this phase (see "Stimuli" section) which, again, were presented sequentially and in a random order. Participants classified each stimulus as an "A" or a "B" by pressing either the "X" or ">" key on the computer keyboard. The allocation of keys to responses was counter-balanced across participants.

In the 1000ms and 2500ms conditions, participants were told that they only had 1 second or 2.5 seconds to make each decision. If they did not respond within this time interval, the stimulus was replaced by the phrase "TIME OUT!" in 2cm high letters. After a five second count-down and a two-second pause, the next stimulus was presented. This time-out procedure was designed to be as salient as possible in order to keep the total number of time-outs low.

Results

Accuracy and mean reaction time data from the no-deadline condition have been reported previously (Wills & McLaren, 1997). All other data are novel.

In the 2500ms condition, 2.87% of trials were timed-out. The figure was 4.84% in the 1000ms condition.

Whilst both rates are relatively low, there were significantly more time-outs in the 1000ms condition, $t(30) = 2.41$, $p < 0.05$. All participants in this experiment made at least sixteen responses before the deadline at each level of stimulus difficulty. Therefore the four slowest responses made by each participant at each level of stimulus difficulty were disregarded in the following analyses (see Introduction for an explanation of this procedure). Time-out trials were counted as the slowest possible responses. For the remaining data, the accuracy, and the mean, standard deviation and skew, for each level of stimulus difficulty and for each participant were calculated.

This data set was subjected to a series of mixed-model ANOVAs, with one within-participants variable (Difficulty, 6 levels) and one between-participants variable (Deadline, 3 levels). A significance level of .05 was set for all analyses. Figures 2 and 3 summarize the data set by providing across-participant averages.

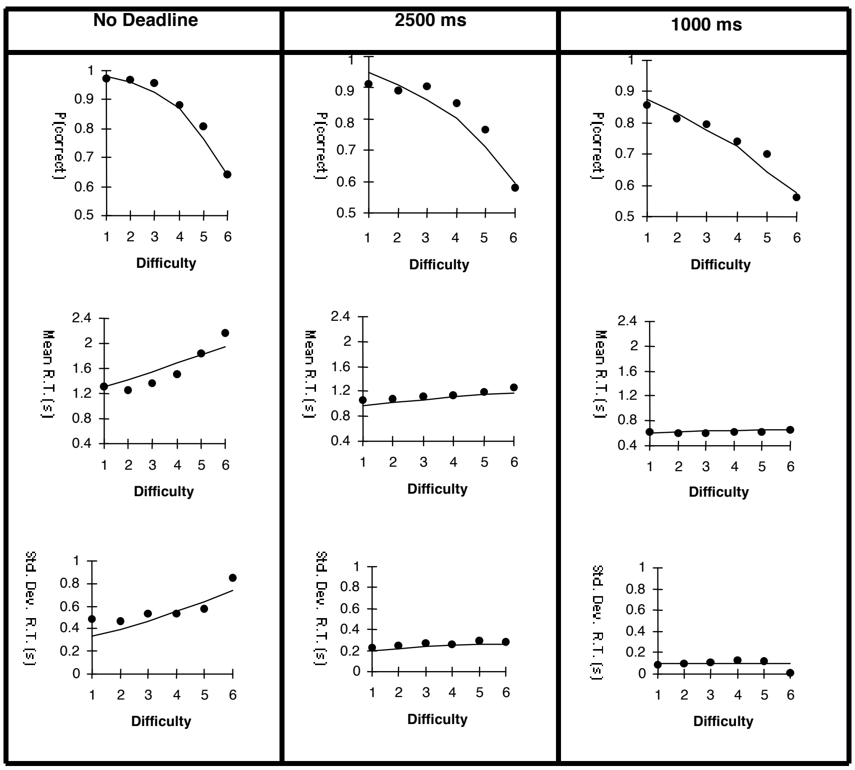

Figure 2: Accuracy, mean reaction time, and standard deviation of reaction time as a function of stimulus difficulty (arbitrary units) and response deadline (milliseconds). Filled markers indicate empirical data. Lines indicate modeling results.

Figure 3 also averages across stimulus difficulty because (as will be seen in a moment) there was no significant effect of stimulus difficulty on skew.

Response accuracy was adversely affected by both stimulus difficulty, $F(5, 205) = 48.1$, and deadline $F(2, 41) = 6.45$. These two factors did not interact significantly, $F(10, 205) < 1$. Mean reaction time increased with stimulus difficulty, $F(5, 205) = 20.19$, and decreased with time pressure, $F(2, 41) = 12.92$. The effect of stimulus difficulty was less pronounced with increasing time pressure, as evidenced by a significant interaction term, $F(10, 205) = 9.32$.

The standard deviation of reaction times increased with stimulus difficulty, $F(5, 205) = 5.77$, and decreased with increasing time pressure, $F(2, 41) = 17.36$. The effect of stimulus difficulty was less pronounced with increasing time pressure, as evidenced by a significant interaction term, $F(10, 205) = 8.45$.

The skew of reaction times decreased with increasing time pressure, $F(2, 41) = 19.52$. However, stimulus difficulty had no significant effect, $F(5, 205) = 1.31$, and the interaction term was non-significant also, $F(10, 205) < 1$. The no-deadline condition shows significantly positive skew, $t(11) = 5.06$, whilst the 1000ms condition shows significantly negative skew, $t(15) = 2.76$. The 2500ms condition showed no significant skew, $t(15) = 0.98$.

Modeling

Wills & McLaren's winner-take-all (WTA) model, like many process models of categorization, assumes that the evidence a presented stimulus is the member of a particular category is represented by a single number or *magnitude term*. In this simulation, the magnitude term for category x (denoted v_x) is $M \times c$, where c is the number of category x elements the presented stimulus contains, and M is a free parameter. Such a relationship sufficiently describes the output of a feature-based, single-layer, delta-rule network taught to classify the stimuli (see Wills & McLaren, 1997 for more details).

The model is illustrated in Figure 4. A single unit represents each category. The magnitude terms for each category are passed to these units as input activation. The output activity of each unit is a function of the total

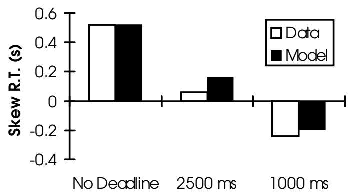

Figure 3: Skew of reaction times as a function of response deadline (milliseconds).

input it receives. Specifically, the output activation of unit i on update c is determined by

$$o_{i,c} = \frac{o_{i,c-1} + En_{i,c}}{1 + En_{i,c} + D} \quad \text{(where } n_{i,c} > 0\text{)} \qquad o_{i,c} = \frac{o_{i,c-1}}{1 - En_{i,c} + D} \quad \text{(where } n_{i,c} \leq 0\text{)}$$

$n_{i,c}$ is the total input to unit i on update c and E and D are constants representing the rate of excitation and decay within the unit.

In addition to the magnitude-term inputs, each unit 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 non-zero activation. Grossberg (1976), and many others since, have employed similar, neurally-inspired decision-making systems.

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

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

where $r_{i,c}$ is the noisy input produced by the magnitude term v_i. The noise in these particular simulations had a range of $+N$ to $-N$, and a rectangular distribution. Superimposed on this noise is the constraint that $r_{i,c}$ cannot exceed one or fall below zero.

The first unit to produce an activation greater than S is assumed to cause the execution of its corresponding response. The number of cycles the unit takes to exceed S represents decision latency, with each cycle representing exactly T seconds.

The model employed includes a number of simplifications, including the assumption that noise is rectangular and that non-decisional components of the categorization process take a fixed T_{res} seconds. Neither simplification is central to the operation of the model - similar predictions can be derived from a model with a variable T_{res} and Gaussian noise. However these simplifications have the advantage of considerably speeding the search of parameter space.

The model described above has seven parameters - *N, E, D, M, S, T* and T_{res}. The basis of the model's predictions is that time pressure reduces the value of S, so S was assigned a different value for each of the three between-participant conditions of the experiment. In all previous applications of the model, it has been assumed that $E = 2D$, and this assumption is continued in the current simulation. T is not a parameter of the model in any important sense, as its only purpose is to convert from one arbitrary unit of time (cycles) to another (seconds). Hence, model fitting involves the manipulation of seven free parameters, from which predictions for 57 data points are to be derived.

Model fitting proceeded via a grid-search procedure. The range and steps of the parameters were N (0.1→3, step 0.1), E (0.01→0.05, step 0.01 and 0.05→0.5, step 0.05), M (0.01→0.08, step 0.01), and S (0.3→0.7, step 0.05) for each of the three S parameters, with the constraint that S did not increase as response deadline decreased. 10,000 decisions were simulated for each permutation of parameters and for each stimulus difficulty level. The cycles-to-decision in each set of 10,000 decisions were then place in rank order, and the 2,000 slowest decisions discarded (in order to mimic the data deletion performed on the empirical data).

This collection of simulated decisions was then employed to produce a set of predictions for each of the permutations of the parameters *N, E, M* and the three S parameters, S_{ND}, S_{2500} and S_{1000}. The relationship between cycles-to-decision and seconds was then estimated for each set of decisions via linear regression

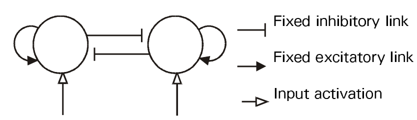

Figure 4: The winner-take-all model

of the 18 empirical mean reaction times to the 18 mean cycles-to-decision. The gradient of the line gives the conversion factor T whilst the intercept provides the value for T_{res}.

Once T and T_{res} were determined for a set of predictions, they were employed to convert each of the 180,000 simulated decision latencies into seconds. Calculations of mean reaction time, reaction time standard deviation and reaction time skew for each stimulus difficulty level in each of the three conditions were then performed using standard formulae.

Scaled root mean square deviations (SRMSD) were used to assess closeness of fit. Scaling was performed by multiplying each empirical data point and each prediction by a factor s. SRMSD was calculated separately for accuracy, mean RT, RT standard deviation, and RT skew predictions. The scaling factors employed were 2, 0.5, 1 and 1 respectively. Total SRMSD was taken to be the sum of these four SRMSDs. The set of parameters providing the best overall fit were as follows N: 2.6, E: 0.03, M: 0.04, S_{ND}: 0.55, S_{2500}: 0.50, S_{1000}: 0.40, T_{res}: 0.033. This is the fit shown in Figures 2 and 3. One cycle of the model was estimated by linear regression to be 0.014 of a second. Cycles-to-decision predicted over 95% of the variance of mean reaction times in this regression (r^2 = 0.953). The SRMSD for accuracy predictions was 0.061, for mean reaction time it was 0.047, for reaction time standard deviation it was 0.062 and for reaction time skew it was 0.063.

Discussion

Imposition of a response deadline decreased the mean, standard deviation and skew of reaction times in a categorization task. From this information about the distribution, one can conclude that these participants adapted to the response deadline in a selective, non-linear manner (as defined in the Introduction). This is a result which, if found to be general, would need to be accommodated by formal models of categorical decisions. The fact that one of the reaction time distributions to be fit is negatively skewed might be considered as a particular source of concern, as categorization have almost uniformly been fit to distributions with some degree of positive skew in the past.

In the space available, it was not possible to evaluate whether all current models of categorical decision have the potential to accommodate the results found. Instead, it was demonstrated that one particular model of categorical decisions (Wills & McLaren, 1997) can mimic the pattern of results found. Wills & McLaren's model is (in approximate terms) a connectionist implementation of a random-walk process (e.g. Laming, 1968). As such, its follows the same basic principles as a variety of other accounts of categorical decision, including stochastic forms of decision-bound theory (Ashby, 2000), extensions of EGCM that can model reaction times (Lamberts, 2000), and Nosofsky & Palmeri's (1997) exemplar-based random walk model. It therefore seems likely that many contemporary models of categorization are capable of accounting for the sort of adaptation to a response deadline observed in this study.

References

Ashby, F. G. (2000). A stochastic version of general recognition theory. *J. Math. Psych., 44*, 310-329.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *JEP: LMC , 14*(1), 33-53.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2,* 50-55.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding. *Biological Cybernetics, 23*, 121-134.

Jones, F. W., Wills, A. J., & McLaren, I. P. L. (1998). Perceptual categorization: Connectionist modelling and decision rules. *Quart. J. Exp. Psy., 51B*(3), 33-58.

Lamberts, K. (1995). Categorization under time pressure. *JEP: General, 124*(2), 161-180.

Lamberts, K. (1998). The time course of categorization. *JEP: LMC, 24*(3), 695-711.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review., 107*(2), 227-260.

Lamberts, K., & Brockdorff, N. (1997). Fast categorization of stimuli with multivalued dimensions. *Memory & Cognition, 25*(3), 296-304.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability and the identification-speed classification relationship. *JEP: HPP, 22*(4), 795-817.

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Percept. & Psychophys., 60*(4), 620-637.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorisation relationship. *JEP: Genral, 115*(1), 39-57.

Nosofksy, R. M., & Palmeri, T. J. (1997a). An exemplar-based random walk model of speeded classification. *Psych. Review, 104(*2), 266-300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*(1), 53-79.

Palmeri, T. J., & Blalock, C. (2000). The role of background knowledge in speeded perceptual categorization. *Cognition, 77*, B45-B57.

Smith, J. D., & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *JEP: General, 113*, 137-159.

van Zandt, T., Colonius, H., & Proctor, R. (2000). A comparison of two response time models applied to perceptual matching. *Psych. Bull. & Rev., 7*(2), 208-256.

Wills, A. J., & McLaren, I. P. L. (1997). Generalization in human category learning. *Quart. J. Exp. Psy., 50A*(3), 607-630.

Acknowledgments

Financial support for this research was provided by Emmanuel College, Cambridge, and by the ESRC. Thanks to Ian McLaren, Stephen Monsell, Stian Reimers, Thomas Palmeri and Koen Lamberts for helpful comments.