Author(s) Stewart, Neil and Brown, Gordon D. A.
Article Title: Sequence Effects in the Categorization of Tones Varying in Frequency
Year of publication: 2004
Link to published version: http://dx.doi.org/10.1037/0278-7393.30.2.416

Running Head: SEQUENCE EFFECTS

Sequence Effects in the Categorization of Tones Varying in Frequency

Neil Stewart

Gordon D. A. Brown

University of Warwick, England

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in categorizing tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 416-430.

Abstract

In contrast to exemplar and decision bound categorization models, the memory and contrast models described here do not assume that long-term representations of stimulus magnitudes are available. Instead, stimuli are assumed to be categorized using only their differences from a few recent stimuli. To test this alternative, sequential effects were examined in a binary categorization of 10 tones varying in frequency. Stimuli up to two trials back in the sequence had a significant effect on the response to the current stimulus. Further, the effects of previous stimuli interacted with one another. A memory and contrast model, according to which only ordinal information about the differences between the current stimulus and recent preceding stimuli is utilized, best accounted for these data.

Sequence Effects in the Categorization of Tones Varying in Frequency

Exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1986) and decision-bound models (e.g., Ashby & Townsend, 1986) are arguably the most successful models of perceptual categorization. These models have a common representational assumption: They assume that stimuli can be represented as points or probability distributions within a multidimensional psychological space. Identification and categorization decisions are then based on these representations. Implicit in this assumption is the notion that the absolute magnitudes (on various psychological dimensions, e.g., loudness, brightness, size) of previously encountered stimuli are available when classifying new stimuli.

There is, however, some evidence to suggest that absolute magnitudes may be unavailable in the identification and classification of simple perceptual stimuli. For example, in a series of classic experiments by Garner (1954), participants' judgments of whether comparison tones were more or less than half as loud as a given reference tone were completely determined by the range of the comparison tones (see also Helson, 1964). Baird, Green, and Luce (1980) demonstrated that two-thirds of the variability in loudness estimates was explained by the variability in the previous estimate when loudnesses were similar, suggesting that the previous loudness is used as a reference point. Such context effects should not be evident if participants did have access to absolute magnitude information. Laming (1997) provided extensive discussion of these and other similar findings

Stewart, Brown, and Chater (2002) suggested that, if absolute magnitude information is not readily accessible, relative magnitude information might instead provide the basis for categorization. Indeed, in the absence of absolute magnitude information the only possible strategy, apart from guessing, is to classify exemplars on the basis of their difference from previous exemplars. Stewart et al. proposed a memory and contrast (hereafter MAC) model of binary categorization, according to which a new exemplar is classified into the same category as an immediately preceding exemplar if it is similar to that exemplar. Alternatively, if the

preceding and current stimuli differ sufficiently, the current exemplar is classified into the alternative category.

The Category Contrast Effect

Stewart et al. (2002) provided some experimental evidence that discriminated between their MAC account and standard exemplar accounts. The paradigm used was a unidimensional binary categorization of ten stimuli, where stimuli of one category took low values on the dimension and stimuli of the other category took high values (see Figure 1). Stewart et al. found that classification of a borderline stimulus (e.g., 5) was more accurate when presentation of the borderline stimulus was preceded by a distant member of the opposite category (e.g., 10) than when it was preceded by a distant member of the same category (e.g. 1). They called this effect the *category contrast* effect.

Stewart et al. (2002) showed that existing exemplar models predict the opposite result. In exemplar models the probability of responding with a given category label is given by the summed similarity of a target exemplar to members of that category, divided by the summed similarity of the target exemplar to members of all competing categories (i.e., in accordance with Luce's, 1959, choice model). If the plausible assumption that stimuli on recent trials are weighted more heavily than those on less recent trials is made (e.g., Nosofsky & Palmeri, 1997), the effect of the immediately preceding exemplar, no matter how dissimilar to the current exemplar, is to increase the summed similarity of the current exemplar to the previous exemplar's category. Thus, according to an exemplar model, a borderline stimulus should be classified more accurately after a distant member of the same category compared to a distant member of the opposite category - the opposite of the category contrast effect.

A MAC strategy does, however, predict the category contrast effect. The similarity between a borderline stimulus and a distant stimulus, either from the same category or the opposite category, is low. Thus, when a borderline stimulus is preceded by a distant stimulus the memory and contrast strategy predicts responding with the opposite category label to that

of the previous stimulus: When the previous stimulus is from the same category, accuracy for responding to the current stimulus will be low, and when the previous stimulus is from the opposite category accuracy for responding to the current stimulus will be high - consistent with the category contrast effect.

Sequential Effects

The purpose of this article is to examine how a MAC strategy might be generalized to include information from recent stimuli other than the preceding stimulus. There is good evidence to suggest that sequential effects in categorization and related tasks are not limited to the immediately preceding stimulus. Stimuli further back in the sequence also have an effect. For example, in absolute identification tasks stimuli that vary along a unidimensional psychological continuum (e.g., the loudness of a tone or the length of a line) are each associated with a unique label. Normally labels are stimulus ranks. In identifying a current stimulus there is a strong assimilative effect to the immediately preceding stimulus (Garner, 1953; Holland & Lockhead, 1968; Hu, 1997; Lacouture, 1997; Lockhead, 1984; Luce, Nosofsky, Green, & Smith, 1982; Mori, 1989; Mori & Ward, 1995; Purks, Callahan, Braida, & Durlach, 1980; Staddon, King, & Lockhead, 1980; Ward & Lockhead, 1970, 1971). In other words, participants are systematically biased to judge the current stimulus as nearer to the previous stimulus than it really is. The effect of stimuli further back in the sequence is the opposite: that is, there is a contrast effect (Holland & Lockhead, 1968; Lacouture, 1997; Ward & Lockhead, 1970, 1971). The contrast effect is smaller in magnitude, and decreases for less recent stimuli, but can be observed for up to the previous five or six stimuli (but see Jestead, Luce, & Green,1977, for an argument that such effects are not direct, but propagated to the current trial through successive responses).

Furthermore, these sequence effects are not due to drifting in responding. Petzold and Haubensak (2001) examined sequential effects in a categorization task with five categories. Stimuli were squares varying in size. There was a significant correlation between stimuli and

responses up to six trials back. These correlations were compared to the expected size of pseudo-sequential effects caused by individual participants showing a drift in their use of the response scale across the experiment. The sequential effects were significantly greater than the expected pseudo-sequential effects up to a lag of two for category judgments.

In summary, in psychophysical tasks there is good evidence that it is not just the immediately preceding stimulus that affects the response on the current trial, but also stimuli further back in the sequence too. If MAC models are to become viable models of categorization then they must offer an account of these sequential effects. We begin by considering how MAC models might be developed to account for the effects of many preceding trials and not just the immediately preceding trial. We then present the predictions of these models. Finally, we present data from a new categorization experiment and use these data to test the models that we have proposed.

Alternative Models

In presenting the alternative models we begin by briefly outlining the original MAC model. We then present some motivation and discussion of the key assumptions in the models we propose before presenting the models and their predictions. We refer to the current trial as trial n, the previous trials as trial n-1, and the kth most recent trial as trial n-k. The physical magnitude of the stimulus on trial n-k is denoted X_{n-k}, the psychological magnitude S_{n-k}, the response R_{n-k}, and the feedback F_{n-k}.

The Original MAC Model

According to the original MAC model proposed by Stewart et al. (2002), participants are assumed to base their categorization decision for S_n on F_{n-1} and on the difference d between S_n and S_{n-1}. In some cases the sign of the difference is sufficient to determine the response. For example, consider the category structure in Figure 1. If F_{n-1} is Category A, and $S_{n-1} \geq S_n$ then S_n must also belong to Category A. This is the case whenever $S_n \leq S_{n-1} \leq 5$ (or when $S_n \geq S_{n-1} \geq 6$ for Category B). When the sign of d is not sufficient to determine the

category of the S_n, then the magnitude of d is used to generate a probability of repeating the F_{n-1} as R_n. When d is small the probability of repeating the label is high and when d is large the probability of repeating the label is small. Originally a Gaussian function was used to relate the distance d to the probability of responding on the current trial with the category label (i.e., feedback) from the previous trial.

$$P(R_n = F_{n-1}) = e^{-cd^2} \quad (1)$$

The free parameter c determines the size of the difference required to give a change in category label by determining how quickly the probability of repeating the previous category label decreases as the difference between the previous and current stimuli increases. Despite making no use of absolute magnitude information, and despite relating the current stimulus only to the immediately preceding stimulus and feedback, such a model can do surprisingly well, achieving, for example, around 85% correct classification performance in binary classification of 10 stimuli (Stewart et al., 2002).

No Long-Term Memory for Absolute Magnitudes

Consistent with the original MAC model (Stewart et al., 2002), we assume that long-term memory traces representing the absolute magnitudes of stimulus attributes are unavailable, or at least unused. We make this strong assumption primarily to demonstrate that models without long-term representation of absolute magnitude information naturally predict the sequential effects observed. This theoretical standpoint is in direct contrast to the representational assumptions of exemplar and decision-bound models of perceptual categorization. Thus our purpose is not to claim that absolute magnitude information is never available or used. Rather, our aim is to show that a natural account of sequence effects in binary categorization of unidimensional stimuli can be provided without such an assumption. In the models that we describe below we assume that absolute magnitude information is available for, at most, a few recent stimuli. An alternative possibility is that this absolute magnitude information is available for only the immediately preceding stimulus, and that

differences between the current stimuli and earlier stimuli are deduced by summing intervening consecutive differences. We shall return to this possibility in the discussion.

Quality of Difference Information

A key theoretical question concerns the nature of the difference information used by participants. Stewart et al. (2002) assumed that the sign and magnitude of the difference S_n - S_{n-1} is available. (If the psychological percept of a magnitude is related to the physical magnitude by a logarithmic transformation, as in Fechner's law, then this assumption corresponds to assuming that only the ratio of the physical magnitudes X_n / X_{n-1} is available.) An alternative and stronger assumption is that only the sign of the difference between successive stimuli is available, and not the magnitude of the difference. In other words, participants are only able to make ordinal judgments, judging whether S_n is greater than, approximately equal to, or smaller than S_{n-1}. Laming (1984, 1997) made a similar claim, and suggested that two additional judgments could be made - 'much greater than' and 'much smaller than.' Laming argued that the assumption that one can only make such judgments is sufficient to account for many of the key phenomena in psychophysics. Here we consider both models where only the sign information is used and models where both the sign and the magnitude of stimulus differences are used.

In some cases, if the sign of the difference between S_n and S_{n-1} is known the category of the current stimulus can be determined from the category of the previous stimulus. Using only the sign of the difference between stimuli might seem unlikely to lead to high levels of performance. However, if one only has access to the sign of the difference between S_n and S_{n-1} (not the magnitude) and F_{n-1} then this strategy does predict a perform above chance (50%) on random sequences of stimuli in a binary categorization with a single category bound (e.g., the category structure in Figure 1). For large set sizes ($N > 10$), average accuracy is about 63% correct, although this performance rises to 69% correct for smaller set sizes ($N = 4$). Here we extend this idea: In general, for a binary categorization with a single category boundary b, we

shall call S_{n-k} *sign-useful* whenever $S_n \leq S_{n-k} < b$ or $S_n \geq S_{n-k} > b$. Note that a participant need know only the sign of the difference between S_n and S_{n-k}, the feedback from trial *n-k*, and the ordering of the categories on the dimension for S_{n-k} to be sign-useful. Knowledge of the category bound is not necessary.

Selection and Integration of Information from Previous Trials

A separate theoretical issue is how information from several previous trials might be selected or combined to inform responding on trial *n*. One possibility is that information from several previous trials is independently combined; a second possibility is that only some of this information is selected and used.

The first possibility can be excluded on the grounds that it does not lead to improved overall accuracy for any model. The argument proceeds as follows. Assume that the proportion of correct responses on trial *n* is given by a weighted sum of the independent contributions of comparisons with *K* previous trials:

$$w_1 \mathrm{f}(S_{n-1}) + w_2 \mathrm{f}(S_{n-2}) + \ldots + w_K \mathrm{f}(S_{n-K}) \qquad (2)$$

where w_k is the weight for trial *n-k*, $\sum_{k=1}^{K} w_k = 1$, and f(S_{n-k}) is the accuracy on trial *n* if only S_{n-k} is used. Averaging over all possible values of each previous stimulus gives an average which reduces to

$$\sum_{S=S^1}^{S^N} \mathrm{f}(S) \qquad (3)$$

where S^i is the psychological magnitude of the *i*th of *N* stimuli. This sum is independent of the weighting used. Thus, only weighting S_{n-1} produces the same overall accuracy as including information from more previous stimuli. In other words, when the effects of information from previous trials are independent of one another, including information from previous trials does not increase the overall accuracy of categorization. Intuitively, this can be understood as follows. When a given previous stimulus is either particularly helpful (or particularly

unhelpful) in classifying the current stimulus, the advantage (or disadvantage) it gives will be diluted when combined with information from other previous stimuli.

The preceding argument rules out any model where the weights allocated to information from previous trials are independent from the stimuli on those trials. An alternative possibility is that the attention paid to a given previous trial will depend on the usefulness of that previous trial relative to other previous trials. For example, comparison might be made to S_{n-3} when and only when comparisons to S_{n-1} and S_{n-2} are not useful. Some evidence consistent with this possibility is provided by Petzold and Haubensak (2001). In their examination of sequential effects described above they obtained an interaction between the effects of the two previous trials. Specifically, the correlation between R_n and R_{n-1} was lower when S_{n-2}'s magnitude was located in between S_n and S_{n-1}'s. Similarly, the correlation between R_n and R_{n-2} was lower when S_{n-1}'s magnitude was located between S_n and S_{n-2}'s. This suggests that when S_{n-1} is nearer S_n than S_{n-2}, S_{n-2} is relied on less and when S_{n-2} is nearer S_n than S_{n-1}, S_{n-1} is relied on less.

In the light of these considerations, in the models we present below we assume that there is a context-dependent selection of the use of stimulus information from particular previous trials.

Availability of Previous Stimuli

In extending the MAC account, we assume that each previous stimulus S_{n-k} ($k > 0$) and corresponding category label can be utilized in categorizing the current stimulus S_n with probability p_{n-k}. For simplicity, these events are assumed to be independent of one another (although this is not a core assumption). Once k is larger than 3 or 4, it is assumed that the probabilities become very small. This corresponds to our assumption that only short-term representations of the absolute magnitudes of stimuli are available, from, at most, a few trials ago. Thus, on each trial, one of a series of possible *states* will occur, with each state corresponding to the pattern of availability or otherwise of previous stimuli. In general, if any

of the past K stimuli can be recalled there will be 2^K states. This assumption forms the basis for all of the MAC models that we present below.

Table 1A provides an example when up to three previous stimuli can be utilized. The top row represents the state where no previous stimuli are recalled. The second row represents the state where S_{n-1} cannot be recalled, S_{n-2} cannot be recalled, and S_{n-3} is recalled. The final row represents the state with probability $p_1\ p_2\ p_3$ where all previous stimuli (up to S_{n-3}) can be recalled.

The Models[1]

We are now in a position to describe three variants of the MAC model. In all of the models that we present, we assume that, if a sign-useful stimulus can be recalled, the current stimulus is correctly classified. Consider the example in Table 1B, when the sequence of stimuli, from S_{n-3} to S_n, is {Stimulus 2, Stimulus 5, Stimulus 7, Stimulus 3}. For the category structure in Figure 1, when classifying Stimulus 3, any of Stimuli 3, 4, or 5 are sign-useful. Thus, whenever $S_{n-2} = 5$ is recalled, as it is in States 3, 4, 7, and 8 (numbering rows from top to bottom), then we assume that participants will respond with the correct R_n.

The models differ only in what happens if no sign-useful stimulus is available. We present four models. If, in a given state, no sign-useful stimulus is recalled then: (a) R_n is guessed (*Guessing Model*); (b) F_{n-1} is given as R_n with probability p_{same} (*Feedback Repetition Model*); (c) the feedback from the last recalled stimulus is given with probability p_{same} or if no stimulus can be recalled, R_n is guessed (*Recalled Stimulus Model*); and (d) the magnitude of the difference between S_n and the last recalled stimulus is used to generate a response according to Equation 1 or if no stimulus can be recalled, R_n is guessed (*Sign and Magnitude Model*). The original MAC model is a special case of model (d) if only S_{n-1} is assumed to be available. The first three models are Sign-Only Models, and the last model is a Sign and Magnitude Model.

Continuing the example in Table 1B, according to the Guessing Model, R_n will be

guessed in States 1, 2, 5, and 6. According to the Feedback Repetition Model, as F_{n-1} = Category B, Category B will be given as R_n in States 1, 2, 5, and 6 with probability p_{same}. When p_{same} = .5 each category is equally likely. Thus, the Guessing Model is a special case of the Feedback Repetition Model. The Recalled Stimulus Model predicts that the response will be guessed in State 1, that in State 2 Category A (the category of the last remembered stimulus S_{n-3} = 2) will be given as R_n with probability p_{same}, and that in States 5 and 6 Category A (the category of the last remembered stimulus S_{n-1} = 7) will be given as R_n with probability p_{same}. In the remaining states, States 3, 4, 7 and 8, one of the stimuli is sign-useful and the correct response is given, as described in the previous paragraph. The Sign and Magnitude Model predicts that the response will be guessed in State 1, in State 2 Category A will be given as R_n with probability $e^{-c(3-2)^2}$, and in States 5 and 6 Category B will be given as R_n with probability $e^{-c(3-7)^2}$. In the remaining states, States 3, 4, 7 and 8, one of the stimuli is sign-useful and the correct response is given.

Model Predictions

Model predictions were generated for each of the models outlined above. For a given model and a given sequence of stimuli, the probability of a correct response can be obtained in three steps. First, the probability of each possible state is calculated. Second, the probability that the model would produce the correct answer in each state is calculated. Third, the probability of being correct is obtained by multiplying the probability of being correct in a state by the probability of that state and then summing over all states. Figures 2-4 show the predictions of the Feedback Repetition Model, the Recalled Stimulus Model, and the Sign and Magnitude Model for the category structure illustrated in Figure 1. (As the Guessing Model is a special case of the Feedback Repetition Model its predictions are omitted.) In these simulations, we assumed that only the previous three stimuli could be recalled, with probabilities p_1 = .9, p_2 = .6 and p_3 = .3. We assumed that p_{same} = .4 for the Feedback

Repetition Model and the Recalled Stimulus Model, and $c = 0.5$ for the Sign and Magnitude Model.

The predictions for the Feedback Repetition Model are shown in Figure 2. The probability of a correct R_n as a function of S_{n-1} (averaging across all possible earlier stimuli) is shown for different S_n. (Lines for values of $S_n > 5$ have been omitted for clarity, but can be generated by reflecting the figure about the line $S_{n-1} = 5.5$.) When $S_n \leq S_{n-1} \leq 5$ (or $S_n \geq S_{n-1} \geq 6$) then S_{n-1} is sign-useful. These cases are represented by the high levels of performance above a .9 chance of being correct. (The probability of being correct is less than 1.0 because S_{n-1} is only available 90% of the time. The probability of being correct is greater than .9 because when S_{n-1} is not available, using earlier stimuli does not always results in an error.) The remaining points correspond to cases where S_{n-1} is not sign-useful. In these cases, S_{n-2} or S_{n-3} may be available (with respective chances of .6 and .3). The probability that one of these earlier stimuli is useful depends on S_n (but not S_{n-1}): When S_n is small (e.g., 1) then there are many possible values of S_{n-2} or S_{n-3} that might be useful (e.g., 2-5); when S_n is larger (e.g., 5) then there are fewer values of S_{n-2} or S_{n-3} that are useful (e.g., only 5). The dependency on the value of S_n as to whether a previous stimulus is useful gives the spreading of the lines in the figure, with better performance for extreme *Sn*.

When S_{n-1} is not useful and if the $p_{same} = .5$ (i.e., R_n is guessed if no sign-useful stimulus can be recalled as in the Guessing Model), then the probability of being correct does not depend on the value of S_{n-1}. When S_{n-1} is not sign-useful and $p_{same} > .5$ then the probability of a correct response is higher if S_{n-1} is from the same category as S_n. When $p_{same} < .5$, then the probability of being correct is higher if S_{n-1} is from the opposite category to S_n. In this case, the category contrast effect is predicted. Accuracy is also higher when $p_{same} < .5$. The optimal value of $p_{same} = .0$. In other words, if no previous stimulus is sign-useful, the best thing to do is give the opposite of F_{n-1} as R_n. This is because when no previous stimulus is sign-useful the correct answer is most often the opposite of the feedback from the previous trial.

In Figure 2B performance is plotted as a function of S_{n-2} (rather than S_{n-1} as in Figure 2A). S_{n-2} has a smaller effect than S_{n-1} because S_{n-2} is available less often. When S_{n-2} is not sign-useful, then the probability of a correct response does not depend on S_{n-2}. Hence the probability of being correct when $S_{n-2} < S_{n-1}$ is the same as when $S_{n-2} > 5$.

In Figure 2C an example is provided for $S_n = 4$ to show the interaction between S_{n-1} and S_{n-2}. If one averages over S_{n-2} (i.e., collapses the plane onto the S_{n-1} axis) then one obtains the $S_n = 4$ line in Figure 2A. If one averages over S_{n-1} one obtains the $S_n = 4$ line in Figure 2B. The ridge labeled A represents predicted high accuracy performance when S_{n-1} is sign-useful (i.e., $S_{n-1} = 4$ or 5). The ridge labeled B represents the case when S_{n-2} is sign-useful. This ridge is of lower accuracy than ridge A because S_{n-2} is assumed to be available less often than S_{n-1}. Where the two ridges intersect and both S_{n-1} and S_{n-2} are useful, it is very likely that at least one will be recalled, and accuracy is very high. The plateaus labeled C-F represent cases when neither S_{n-1} or S_{n-2} is sign-useful. For plateaus C and D, S_{n-1} is from the opposite category to S_n and so the correct answer is to give the opposite of F_{n-1} as R_n. As $p_{same} = .4$ the correct answer is given 1 - .4 = .6 of the time. For plateaus E and F, S_{n-1} is from the same category as S_n and so the correct answer is to give F_{n-1} as R_n. Thus the correct answer is given .4 of the time.

The Feedback Repetition and Recalled Stimulus models make very similar predictions. The only difference between the two models is that if no sign-useful previous stimulus can be remembered in the Feedback Repetition Model then F_{n-1} is repeated with some probability and in the Recalled Stimulus Model then the feedback to the last remembered stimulus is repeated with some probability. Given that the previous stimulus is very likely to have been recalled, in practice the two models are almost equivalent. The only difference is that S_{n-2} does have a small effect when no useful stimulus can be recalled in the Recalled Stimulus Model (i.e., the horizontal components of Figure 3B are not constrained to be of equal value on the left and right of the figure, as in Figure 2B). Apart from this small difference, the properties of the predictions do not differ between the two models. For this reason, we do not consider the

Recalled Stimulus Model further and instead will focus on the simpler Feedback Repetition Model.

In the Sign and Magnitude Model, the signs and magnitudes of differences between the current stimulus and recent previous stimuli are available, rather than just the signs of the differences as in the other models. If none of the previous stimuli that are recalled are sign-useful, then the difference from the last remembered stimulus is used. The smaller the difference, the more likely that the category label feedback from the last stimulus is given in response; the bigger the difference the more likely that the response will be the category that was not indicated by the previous trial's feedback. Figure 4 shows the predictions of the Sign and Magnitude Model with the same availability of previous stimuli as in the previous two simulations.

Figure 4A plots the probability of a correct R_n as a function of S_{n-1} (averaging across all possible earlier stimuli) for different S_n. When S_{n-1} is sign-useful (i.e., $S_n \leq S_{n-1} \leq 5$ or $S_n \geq S_{n-1} \geq 6$) then performance is high, as in the Sign Only Models. When S_{n-1} is not useful, then accuracy depends on the difference between S_{n-1} and S_n. When S_{n-1} is similar to S_n, then the probability of giving F_{n-1} as R_n is high. Thus, when $S_{n-1} = 6$ and $S_n = 5$, this will lead to errors as the two stimuli come from opposite categories. As S_{n-1} increases above 6, the difference grows and the model predicts that swapping, which is the correct response, is more likely. When $S_{n-1} = 4$ and $S_n = 5$, then, as the two stimuli are similar, F_{n-1} is repeated as R_n, leading to accurate performance. As S_{n-1} decreases below 4, the difference increases, and swapping, which is incorrect, becomes more likely. The Sign and Magnitude Model thus necessarily predicts the category contrast effect. Figure 4B shows the same qualitative pattern of performance as a function of S_{n-2} not S_{n-1}. The pattern is greatly attenuated, as S_{n-2} is assumed to be available less often than S_{n-1}.

Figure 4C is analogous to Figures 2C and 3C. In Figures 2C and 3C, when S_{n-1} and S_{n-2} are not useful, only the category of the previous stimuli is important (i.e., regions C-F are

plateaus). However, in the Sign and Magnitude Models, the difference between stimuli, rather than just the category of the previous stimuli, determines responding. Thus, in the same regions, performance is predicted to vary as a function of the magnitudes of S_{n-1} and S_{n-2} (or, more specifically, the differences from S_n, but S_n is held constant at 4 in this plot). The explanation of this pattern is as above.

Experiment

In the previous modeling section, we outlined two categories of model. In one model, only the sign of the difference between previous stimuli was available. In the other model, both the sign and the difference of the previous stimulus was available. These two classes of models necessarily make quite different predictions. The Sign Only Models predict that, when no previous stimulus is sign-useful, only the category of previous stimuli will influence responding on the current trial. The Sign and Difference Models predict that, when no previous stimulus is sign-useful, the magnitude of the previous stimuli will have a continuous effect. The purpose of this experiment is to test these two contrasting accounts. For this purpose, participants classified a truly random sequence of stimuli to examine the effects of previous stimuli on classification of the current stimulus.

Method

Participants. Sixteen undergraduate psychology students volunteered to participate.

Stimuli. Ten 500-ms sine-wave tones of differing frequency were used as stimuli in this experiment. Each tone was 6% higher in frequency than the tone immediately lower in frequency, and thus the tones were equally spaced on a log-frequency scale. The first tone had a frequency of 768.70 Hz, and the last tone had a frequency of 1298.70 Hz. The 10 tones were divided into two categories, with the 5 lowest frequency tones in one category, and the 5 highest frequency tones in the other category. The amplitude of stimuli was linearly ramped from zero to maximum in the first 50 ms of the stimulus and from maximum to zero in the last 50 ms of the stimulus to prevent click artifacts at the stimulus onset and offset. Stimuli were

transduced using a Creative Labs Ensoniq CT5880 audio PCI sound card and Sennheiser eH2270 headphones.

Procedure. Participants were tested one at a time in a quiet room. Participants were informed that they would hear a number of tones, one after the other. They were told that low tones belonged to one category and high tones belonged to another, and that after each tone they would be asked to respond with the category they thought the tone came from. Although at first participants would have to guess, they were informed that by attending to the correct answer displayed on the screen after each response, they could learn which tones belonged to which category. They were given an opportunity to ask the experimenter questions before the experiment began.

Each trial began with a tone randomly selected (with replacement) from the set. A '?' prompt appeared on the screen with the onset of the tone. From the onset of the tone participants were able to respond with either 'Z' or 'X' (labeled 'A' and 'B' respectively) on a standard keyboard. The assignment of labels to categories was counterbalanced across participants. The '?' prompt disappeared immediately after participants responded. After the participants had responded or 500 ms after the offset of the tone, whichever was later, the correct answer (either 'A' or 'B') was displayed on the screen for 500 ms. Feedback was given throughout the experiment. There was a 500-ms pause before the next trial began. There were six blocks each of 100 trials. Participants were given a break between each block.

Results

Accuracy. Figure 5 shows the proportion of correct responses as a function of S_n. Performance has been averaged across the two categories, so that the abscissa represents the stimuli furthest from the category bound on the left and those closest on the right. Performance is highest on the extreme stimuli and decreases monotonically to be lowest for the borderline stimuli, $F(4, 60) = 188.93$, $p < .0001$ (one-way univariate ANOVA).

Sequence effects. Figure 6 shows the interaction between current and previous stimuli.

Figure 6A shows the interaction between S_n and S_{n-1}. When S_{n-1} is sign-useful, accuracy is high. Otherwise, accuracy depends only on the category of S_{n-1}, with higher accuracy when S_{n-1} belongs to the opposite category to S_n, consistent with the original category contrast effect. A two-way univariate ANOVA was run, with factors S_n and S_{n-1}. There was a main effect of S_n, $F(4, 52) = 176.10$, $p < .0001$. There was a main effect of S_{n-1}, $F(9, 117) = 3.92$, $p = .0002$. There was also significant interaction, $F(36, 468) = 4.27$, $p < .0001$. Figure 6B shows the interaction between S_n and S_{n-2}. Another two-way univariate ANOVA was run, with factors S_n and S_{n-2}. There was a main effect of S_n, $F(4, 40) = 129.36$, $p < .0001$. There was a main effect of S_{n-2}, $F(9, 90) = 4.42$, $p < .0001$. There was also significant interaction, $F(36, 360) = 1.66$, $p = .0117$.[2] Effects of S_{n-3} and S_{n-4} and their interactions with S_n were examined. Although the pattern of data was similar to that for S_{n-1} and S_{n-2}, the effects and interactions were not significant: largest $F(1, 117) = 1.62$, $p = .1162$. We think that it is likely that significant effects would be found in a more powerful experiment.

The category contrast effect. It is possible to test the category contrast effect observed by Stewart et al. (2002). Recall that the category contrast effect is defined as more accurate categorizations of a borderline stimulus following a distant stimulus from the opposite category compared to a distant stimulus from the same category. This effect can be examined in the current data by comparing accuracy for $S_n = 5$ (the borderline stimulus) in Figure 6A when $S_{n-1} = 10$ (distant stimulus from the opposite category) and when $S_{n-1} = 1$ (the distant stimulus from the same category). A t-test shows that although the direction of the difference is consistent with the category contrast effect, the difference is not significant, one tailed $t(15) = 1.26$, $p = .1143$.[3] However, we have already noted that only the category of S_{n-1} seems to matter. When performance averaged over $1 \leq S_{n-1} \leq 4$ is compared to performance averaged over $7 \leq S_{n-1} \leq 10$, performance is higher when S_{n-1} is from the opposite category, consistent with the category contrast effect, one tailed $t(15) = 2.53$, $p = .0116$. A similar analysis was performed for S_{n-2}. The category contrast effect was significant when only extreme values of

S_{n-2} were used in the comparison, one tailed $t(15) = 2.63$, $p = .0094$, and when only the category of S_{n-2} was used, one tailed $t(15) = 2.01$, $p = .0316$.

Conditional sequence effects. A strong prediction of the MAC models described here is that when a previous stimulus is recalled and is sign-useful, other previous stimuli will have no effect.[4] To investigate this prediction was assumed that S_{n-1} is nearly always available. Performance is plotted as a function of S_n and S_{n-2} in Figures 7A and B. Performance when S_{n-1} is sign-useful is plotted in Figure 7A and when S_{n-1} is not sign-useful in Figure 7B. (Note that few data are available for the line $S_n = 5$, as it is quite rare that S_{n-1} is useful when $S_n = 5$: Some of the data points are based on as few as 10 responses, and hence the standard error of the mean is large.) As we are assuming that S_{n-1} is almost always available, S_{n-2} should have no effect when S_{n-1} is sign useful and S_{n-2} should have an effect when S_{n-1} is not sign-useful. This is the pattern that is observed. Figures 7C and D contain the analogous examination of the effect of S_{n-1} when S_{n-2} is sign-useful (Figure 7C) and is not sign-useful (Figure 7D). We have no way of knowing whether S_{n-2} was available to participants on each of these trials, as we have no independent measure of the availability of stimuli (we do not assume that S_{n-2} is nearly always available). Thus, although S_{n-2} was potentially useful on all of the trials that contributed to Figure 7C, it was not always available. Thus we predict an attenuated effect of S_{n-1} in Figure 7C compared to Figure 7D. This is the pattern that is observed.

Discussion

The purpose of this experiment was to examine the effect of previous stimuli on classification of a current stimulus in a binary categorization. We found large effects of previous trials, and interactions between these effects that were consistent with the predictions of Sign-Only MAC models.[5] Specifically, if S_{n-k} was sign-useful (i.e., using only the sign of the difference between S_{n-k} and S_n allowed the category of S_n to be determined) then R_n was accurate and unaffected by other preceding stimuli. An interaction of this sort is entirely consistent with the interaction found by Petzold and Haubensak (2001). When S_{n-k} was not

sign-useful, only the category of S_{n-k} determined responding, as assumed in the Sign-Only MAC models. As we found an effect of the category of S_{n-2}, this rules out Guessing and Feedback Repetition models which do not predict this effect (these models specify that when no sign-useful previous stimulus is available, then either R_n is guessed, or F_{n-1} is repeated with some probability - both predict no effect of S_{n-2}).

The category contrast effect was also replicated: when S_{n-k} was not sign useful, R_n was more accurate when S_{n-k} was from the opposite category. However, we found no evidence that the magnitude of the difference between stimuli was utilized, as was assumed in the original MAC model (Stewart, et al., 2002). It seems that this original explanation of the category contrast effect must be modified. The explanation offered here is that of the Recalled Stimulus MAC model: that is, when no previous stimulus is sign-useful, it is optimal for the feedback from the last recalled stimulus to be repeated with probability $p_{same} < .5$. When this is the case then the Recalled Stimulus MAC model predicts greater accuracy.

Sequential effects have been examined extensively in the magnitude estimation paradigm which was popularized by Stevens. Typically, in multiple regression analyses, current responses are contrasted with previous stimuli and assimilated towards previous responses (e.g., Jestead et al., 1977; Ward, 1982, 1985, 1987). Our category contrast effect is also an example of a contrastive relationship between the previous stimulus and the current response, albeit in a different task. The true nature of the relationship between the previous stimulus and the current response in magnitude estimation tasks may not be contrastive. Instead, the negative coefficient of S_{n-1} revealed in many multiple regression analyses may confound a (possibly additive or assimilative) perceptual effect of S_{n-1} with a hidden autocorrelated error in the judgment process (DeCarlo, 1992, 1994; DeCarlo & Cross, 1990). This possibility goes some way towards reconciling the assimilation seen in absolute identification with the contrast seen in magnitude estimation and in our own categorization task. We return to this possibility below.

Not all authors find a contrastive effect of the previous stimulus. In magnitude estimations of length and numerousness, Morris and Rule (1988) found that the correlation between the deviation from the mean response to a given stimulus on the current trial and the previous stimulus was very small. Morris and Rule accounted for their lack of a contrast effect compared to that found by previous authors by suggesting that the contrastive relationship is a sensory effect that would differ between different stimulus continua (p. 72). This account may well explain the discrepancy between our results and Morris and Rules's. An alternative is to note that in magnitude estimation tasks, to the extent that participants can perform the task, stimulus magnitudes and responses are correlated. As the effects of previous stimuli and responses on the current response are opposite, they may well have canceled out in Morris and Rule's data (for an example of such cancellation, see Schifferstein & Frijters, 1992). Haubensak (1992a) also failed to find an effect of preceding stimuli. Haubensak had participants judge the size of 10 squares using six categories without feedback. To investigate whether contrast effects occurred, Haubensak varied the sequential dependencies. For one group of participants, the probability with which the current stimulus came from the same half of the continuum was .75 and for another group it was .25. Haubensak found that the ratings given to each square did not differ between the two groups. Under the hypothesis of contrast between successive stimuli, ratings should have been higher for the first half of the scale and lower for the second half of the scale for the .75 Group (as in the .75 Group the preceding stimulus more likely to have been from the same half of the continuum). However, a contrastive model only predicts very small differences in this paradigm. For example, if the 10 stimuli are assumed to be evenly spaced psychologically and contrast is set to be 10% of the difference between stimuli, then the mean rating for squares differ by only 0.14 on a six point rating scale between the two groups. Further, although Haubensak found that the previous stimuli did not systematically bias the current response, if previous stimuli are similar, the error in the current response is greatly reduced, at least in absolute identification of loudness (Luce

et al., 1982; Nosofsky, 1983) and length (Hu, 1997).

Jestead et al. (1977) argued that sequential effects in psychophysical tasks such as magnitude estimation, absolute identification, and categorization extend to only the immediately preceding stimulus. They describe the notion that the effects may extend to stimuli further back in the sequence as "appalling" (p. 92) and "disturbing" (p. 93). They argue that the depth of sequential effects observed by Lockhead and his colleagues (e.g., Holland & Lockhead, 1968; Ward & Lockhead, 1970, 1971; Ward, 1972, 1973) are artifactual and that only the immediately preceding influence affects responding on the current trial. They argue that deeper sequential effects are in fact caused by a propagation of errors in responding to subsequent stimuli (see also DeCarlo, 1992, 1994; DeCarlo & Cross, 1990). For example, S_{n-2} and R_{n-2} would affect R_{n-1}, which in turn would affect R_n. S_{n-2} would not directly affect R_n. In the present data, we have found significant sequential effects up to a depth of two previous trials (consistent with Petzold & Haubensak, 2001, though their task was different). However, responses contained only one bit of information (either Category A or B). When S_{n-2} is from the same category as S_n, sometimes S_{n-2} will be sign-useful and sometimes it will not. F_{n-2} or R_{n-2} do not contain this information. Thus it is hard to see how information about S_{n-2} can be propagated to the current trial through consecutive responses. The "appalling" and "disturbing" alternative is that S_{n-2} directly affects R_n, as we assume in extending the MAC model to account for these effects. A second alternative is to postulate an intermediate decision scale that lies between a sensory scale and a response scale, breaking with the traditional form of Stevens (1957) assumption that responses are directly proportional to the sensation magnitude. This might well be the case in our binary categorization task. Sequential effects could then be propagated through successive errors in judgment error on the decision scale that would not be observed on the much coarser binary response scale. However, if this is the case, why is the difference magnitude information that would be available from the postulated decision scale not used in judgment?

General Discussion

In this article we have considered how the MAC strategy might be extended to account for the effects of not only the immediately preceding stimulus, but also other recent stimuli. There are several psychological axioms that lie behind the extension of the model. First, information about the absolute magnitudes of stimulus properties is only available from the short-term memory representations of very recent trials. We have assumed, for the purposes of illustration, that there are no long-term representations of absolute magnitudes. Second, the differences between recent stimuli and the current stimulus are used to classify the current stimulus. We tested two alternatives here: (a) that only the sign of the differences is utilized or (b) that both the sign and magnitude of the differences are utilized. Third, information from each preceding trial is not used independently, but instead there is an interaction, such that whether or not information from one previous trial is used depends on other previous trials. The data that we presented in the Experiment are most consistent with the idea that only the sign of differences between the current stimulus and preceding stimuli is utilized, at least for the present category structure and stimuli, and that this information is used in a context-dependent way. Before closing this article we address some of the issues raised by these ideas.

Relative and Absolute Models of Categorization

Traditional models of categorization, such as exemplar models, assume that the absolute magnitude of a stimulus provides the basis for categorization. MAC models assume that it is the differences between the current stimulus and recent stimuli that provide the basis for categorization. We do not wish to suggest that absolute magnitudes are completely unavailable and unrepresented in the sensory pathways. A simple thought experiment shows that such a position is untenable. Consider categorizing tones on the basis of their frequency. The original claim that only the difference between S_{n-1} and S_n is used in the decision process makes the implicit assumption that the absolute magnitude of S_{n-1} is temporarily represented:

Without some temporary memory of the absolute magnitude of the previous stimulus being maintained over the silent inter-trial interval then it would be impossible to construct or perceive the difference between the current stimulus and the previous stimulus. Constructing such a difference involves comparing S_{n-1} and S_n, and thus some representation of the absolute magnitude of S_{n-1} must be maintained until, and available when, S_n is perceived.[6] The claim that we are making is that the long-term memory representations of stimuli do not contain absolute magnitude information (or that if they do, the information is not used in a binary categorization of tones varying in frequency). Thus, long-term memory representations differ qualitatively from the representations of current or very recent stimuli.

Long-Term Frames of Reference

The data of Ward and Lockhead (1970) and Ward (1987) are problematic for the claim that there is no long-term representation of absolute magnitudes. Stimuli varying in loudness were used in different psychophysical tasks. Some tasks - absolute identification with feedback, absolute identification without feedback (i.e., category judgment), and ratio magnitude estimation of successive stimuli - required participants to make relative judgments (Ward & Lockhead, 1970; Experiments 1 and 2 of Ward, 1987). Other tasks - absolute magnitude estimation and cross modality matching (to duration) - required participants to make absolute judgments (Experiments 3 and 4 of Ward, 1987). For a given task, sessions were repeated over several days. The crucial manipulation was to vary the loudness of the stimulus set on different days, by either increasing or decreasing the intensity of all of the stimuli by a constant. No matter whether the participants were supposed to be making absolute or relative judgments, judgments in a given session were biased by the stimulus-response mapping from the previous day's session. Participants had a tendency to respond to stimuli as if they could partly recall the responses associated with stimuli on the previous day's session. These data are the strongest evidence we know of for long-term frames in psychophysical judgment, which in this case are given by the previous day's stimulus-response

mapping.

How might Ward's data be reconciled with our own? If people can store only a single frequency in the long-term, they should do very well in a binary categorization by storing a stimulus at the category boundary, given that adjacent stimuli are easily discriminable. Thus, we conclude that such a representation is either unavailable or very poor (or "fuzzy", as Ward, 1987, p. 226, suggests).

Our data concern frequencies of tones; Ward's concerned the loudness of tones. Long-term absolute magnitudes may be available only for loudnesses, or unavailable only for frequencies (few people have perfect pitch; about 0.01% of the general population, Takeuchi & Hulse, 1993). An alternative is that a long-term frame of reference is available for frequency, but for some reason is not utilized in our task. Certainly, task characteristics seem to be able to alter the balance between short and long term frames of reference. DeCarlo showed that a long-term frame of reference was more heavily weighted in a regression equation fitted to magnitude estimation data when instructions suggested a long-term frame of reference (DeCarlo & Cross, 1990; DeCarlo, 1994) or when inter-trial intervals were large (DeCarlo, 1992).[7]

A second source of evidence that seems problematic for the claim that there are no long-term frames of reference concerns the ubiquitous bow effect: Typically, performance is better on extreme stimuli in an absolute identification than for central stimuli (Braida & Durlach, 1972; Durlach & Braida, 1969; Lacouture, 1997; Lacouture & Marley, 1995; Luce, Green, & Weber, 1976; Pollack, 1952, 1953; Weber, Green, & Luce, 1977). While the restricted ability to make mistakes at the ends of the range certainly contributes to the bow effect, many authors attribute the effect to differential sensitivity along the stimulus range (Berliner, Durlach, & Braida, 1977; Braida & Durlach, 1972; Luce et al., 1982; Shiffrin & Nosofsky, 1994) or memory for the extreme stimuli (Berliner & Durlach, 1973; Braida et al., 1984; Marley & Cook, 1984; see also Gravetter & Lockhead, 1973). However, the bow effect

can be explained without recourse to long-term frames of reference. Each variant of the MAC model presented here predicts better performance for extreme stimuli, because extreme stimuli are more likely to have a sign-useful stimulus in the set of recent stimuli. This approach can also be applied to absolute identification. Unpublished modeling from our laboratory demonstrates that a bow effect can be produced if the sign of the difference between the previous stimulus and the current stimulus is used to restrict the range of possible responses, from which one is selected at random. This is because the range of possible responses generated in this way always contains the correct response and, on average, is smaller for extreme stimuli. Thus the bow effect need not be interpreted as evidence for long-term frames of reference.

Reconciling Exemplar and MAC Models

It may seem that the MAC models described here are hard to reconcile with exemplar models of categorization, that currently dominate the literature. However, Stewart and Brown (2003) showed that this is not the case. In exemplar models, the similarity between two exemplars is a function of the difference between them, and thus although exemplar models assume that absolute magnitudes are available, classification of stimuli depends only on stimulus differences.

Stewart and Brown (2003) suggested that the issue of the availability of absolute or relative magnitudes may be reconceptualized in terms of the availability of exemplars in memory. Stewart and Brown developed an extension of an exemplar model, the generalized context model (GCM, Nosofsky, 1986), by modifying the choice rule. The key psychological claim instantiated in this extension is that not only is the similarity of a novel exemplar to stored exemplars of a category evidence that the novel exemplar belongs to that category, but also the dissimilarity to alternative categories. This modification allows a large dissimilarity (or a small similarity) to count as evidence against category membership (cf. standard models where very low similarity counts as evidence, albeit very slight, for category membership). The

original MAC model was shown to be a special case when only recent exemplars were assumed to be available. When all exemplars were assumed to be available, Stewart and Brown's extension closely mimics the GCM. Thus, the only real difference between the two accounts is that the MAC model assumes that only immediately preceding stimuli are available, and that long-term memory representations of the absolute magnitudes of stimuli cannot be formed. Conversely, the GCM assumes that every previously encountered exemplar's absolute magnitude is stored in long-term memory.

Direct or Indirect Availability of Differences

In our extension of the MAC model to utilize information from prior trials other than the immediately preceding trial, we assumed that the information about previous trials is directly available. For example, we assume that the difference between S_n and S_{n-2} is deduced by comparing the short-term memory trace of S_{n-2} with S_n. An alternative is that such a comparison is not made because S_{n-2} is not available in memory when S_n is encountered. Instead long-range differences could be deduced by summing intervening consecutive differences. For example the difference between S_{n-2} and S_n might be derived by summing the difference between S_{n-1} and S_{n-2} with the difference between S_n and S_{n-1}. If this is the case, then why is the magnitude of difference information not used (see Experiment)? It would seem unlikely that the magnitudes of differences were used to estimate long-range differences, and yet were not utilized in categorization.[8]

In the Sign-Only MAC Models that receive support from the Experiment, we assume that only ordinal judgments are used in a simple binary categorization. It is possible to infer something about long-range ordinal judgments if only ordinal judgments between consecutive stimuli are available. Table 2 it is shown that the sign of the S_{n-2} - S_n difference can only be deduced from intervening consecutive differences when S_{n-2} - S_{n-1} and S_{n-1} - S_n are of the same sign. In these cases either $S_n \geq S_{n-1} \geq S_{n-2}$ or $S_n \leq S_{n-1} \leq S_{n-2}$. (In the remaining cases the sign of the S_{n-2} - S_n difference can only be deduced if the magnitude of the intervening consecutive

differences is available.) Thus, S_{n-2} can only be sign-useful if its sign is deduced from only ordinal information about intervening consecutive differences, when S_{n-1} is also sign-useful. The data illustrated in Figure 7B show that S_{n-2} is sign-useful even when S_{n-1} is not. Thus we suggest that S_{n-2} is not deduced from only ordinal information about intervening consecutive differences.

To conclude, together the above two arguments suggest that the sign of longer-range stimulus differences is deduced by comparing S_n directly with the short-term memory trace of S_{n-k}. This conclusion necessarily implies that some absolute magnitude information is available, at least in the short term, for preceding stimuli other than just the immediately preceding stimulus.

Relationship to Existing Models

Petzold and Haubensak's (in press) multiple standards model provides an account of how unidimensional stimuli are categorized. The model uses the nearest upper and lower reference points taken from the previous two stimuli and the maximum and minimum stimuli. This selection of reference points is motivated by Haubensak's (1992b) consistency model. Thus, the relative arrangement of S_n, S_{n-1}, and S_{n-2} determines the reference points selected. For example, consider the absolute identification of 10 evenly spaced stimuli. If $S_{n-2} = 6$, $S_{n-1} = 9$, and $S_n = 4$, then $S_{n-2} = 6$ will be selected as the nearest upper anchor S_U and Stimulus 1 will be selected as the lowest anchor S_L. R_n is then generated by linear interpolation (i.e., responding such that $(R_n - R_L) / (R_U - R_L)$ matches the ratio of the stimulus magnitudes $(S_n - S_L) / (S_U - S_L)$). (This covers the case when there are fewer categories than stimuli.) Thus the multiple standards model differs from the MAC model we advocate as it assumes that (a) absolute magnitudes are maintained in long-term memory, at least for the extreme stimuli, and (b) that the magnitudes and signs of the differences are used. However the multiple standards model and the MAC models described here have in common the assumption that the selection of stimuli as standards depends on the magnitudes of those stimuli and the magnitudes of other

recent stimuli.

DeCarlo also suggests that multiple standards might be used in making psychophysical judgments. In his dynamic theory of proportional judgment (DeCarlo & Cross, 1990; DeCarlo, 1992, 1994), two frames of reference are suggested. One frame comprises a long-term reference stimulus (perhaps the first stimulus) and the response assigned to it. The other frame comprises the previous stimulus and the previous response. The relative reliance upon each frame is modulated by a free parameter. Our data presented in the Experiment and those of Petzold and Haubensak (2001) described above suggest that the frame of reference may change on a trial-by-trial basis, depending on the current and immediately preceding stimulus. Such trial-by-trial changes in the frame of reference could, in principle, be accommodated within the dynamic theory of proportional judgment by specifying how the free parameter modulating the relative contributions of each frame might be made to depend on the context provided by current and recent stimuli. For example. DeCarlo and Cross (1990, p. 387) provide an account of the observation that the autocorrelation of successive responses is highest when stimuli are similar (e.g., Baird et al., 1980; Jestead et al., 1977; Schifferstein & Frijters, 1992) by allowing the free parameter to be proportional to the difference between successive stimuli. However, for such an approach to be successful in its application to our data, additional frames of reference would need to be included for S_{n-2} and perhaps further back. This, then, would be a departure from another key aspect of the model which attributes long range sequential dependencies in the data to the propagation of judgment errors through successive responses (as described in the Discussion of the Experiment above).

Treisman and Williams (1984) proposed a theory of criterion setting to account for the sequential effects observed in many psychophysical tasks. The basis of the model is to assume a Thurstonian sensory scale, which is divided into response categories by criteria. This model then immediately differs from MAC models by assuming that absolute magnitudes of stimulus properties are directly available to the decision process. Two opposing, short-term

mechanisms act on the criteria on a trial-by-trial basis. First, a tracking mechanism, motivated by the assumption that objects in the real world tend to persist (Treisman & Williams, 1984, p. 94), moves criteria away from the currently perceived sensory effect. This increases the probability of repeating the response associated with the stimulus that was most likely to have caused that sensory effect. Second, a stabilizing mechanism acts to locate criteria nearer to the prevailing flux of sensory inputs to increase the amount of information transmitted (Treisman & Williams, 1984, p. 94). The stabilizing mechanism acts in opposition to the tracking mechanism, and Treisman and Williams suggest that these two mechanisms are each affected by different variables. The model can, in principle, account for the category contrast effect, if it is assumed that the stabilization mechanism dominates. For example, consider in our experiment, Stimulus 1 followed by Stimulus 5. Stimulus 1 would cause the category bound to move down the scale. Thus, Stimulus 5 is now more likely to be categorized into Category B, as seen in the category contrast effect. If Stimulus 10 precedes Stimulus 5, then the stabilizing mechanism will move the bound up the scale, making a Category A response to Stimulus 5 more likely.

We cannot see a clear theoretical motivation for assuming that stabilization should dominate in an binary categorization task with feedback. Such an argument would have to rest on the task properties specific to binary categorization favoring stabilization over tracking. Treisman and Williams (1984, pp. 103-104) suggest that feedback should reduce tracking as bounds are only modified by tracking when the correct response is given. This argument predicts that as the number of categories is reduced in a categorization task and accuracy rises, tracking should be increased. Thus more assimilation is predicted. Treisman and Williams also suggest that stabilization should increase as the number of categories increases (p. 104). Thus, with few categories there should be little stabilization. In summary, these two arguments predict that tracking should dominate stabilization, which in turn would predict a category assimilation effect.

Further, the criterion-setting model cannot account for the interaction between the effects of previous stimuli seen in Petzold and Haubensak's (2001) data, as Petzold and Haubensak themselves point out (p. 970), or the interactions in our data, as the effect of previous trials is accumulated in a linear sum (Treisman & Williams, 1984, p. 75). Despite this failure, the generality of the criterion-setting model suggests that it should be pursued as an account of sequential effects.

In applying the judgment option model (a component of complementarity theory) to category judgment, Baird (1997) classifies contextual effects as a product of the response process, as we do in our MAC models. Specifically, his model assumes that variability in psychophysical judgment can be attributed to uncertainty about which response should be associated with a given stimulus. Baird's model differs from the MAC models in that absolute magnitudes of stimuli are assumed to be available. Participants are assumed to use stimulus-response pairings, which, following Haubsensak (1992b), are determined by the responses early in the experimental session. Subsequently, the rank-order decision rule is used within the judgment option model, whereby participants give responses that follow the same rank order as the stimuli. Baird discusses the example for a binary categorization (p. 209), giving three explicit rules: (a) if $S_n = S_{n-1}$, $R_n = R_{n-1}$, (b) if $S_n > S_{n-1}$, R_n = Catgeory B, (c) if $S_n < S_n$-1, R_n = Category A. These rules bear some relation to the concept of sign-usefulness we introduced earlier. However, Baird's rules are more generally applicable. Our data do not support this generality. For example, (b) suggests that if Stimulus 1 is judged as Category A, then any larger subsequent stimulus should be judged Category B. Figure 6 shows that this is not the case. When Stimuli 2-5 follow Stimulus 1, they are more often categorized into Category A. Although Baird's complementarity theory is considerably more general than our MAC models, it does not offer an obvious account of the sequential effects in our data.

Laming (1997) argues that the judgments underlying performance in many psychophysical tasks are no better than ordinal (i.e., 'greater than', 'the same as', and 'less

than'). This argument was motivated by the observation that the variability in performance in magnitude estimation, absolute identification, and cross-modal matching is typically several orders of magnitude greater than that which is associated with discrimination of the same stimuli. Laming then demonstrates that ordinal judgments of this kind are sufficient to explain much of the data from psychophysical experiments, to explain the dependency of the results on the range of stimuli selected for the experiment, and to explain the sequential effects observed in these tasks. The data we present in this Experiment are consistent with Laming's assumption that judgments are ordinal, and this assumption is at the core of the MAC model that we present to account for the data. Where we differ from Laming is in assuming that ordinal judgments are not just possible for the comparison of the immediately preceding stimulus with the current stimulus, but also for other very recent stimuli. Our MAC model is consistent with Laming's key claim: that there is no internal scale of sensation and that instead, judgment is relative to the context in which stimuli are presented.

References

Abeles, M., & Goldstein, M. H. (1970). Functional architecture in cat primary auditory cortex: Columnar organization and organization according to depth. *Journal of Neurophysiology, 33,* 172-187.

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

Baird, J. C., Green, D. M., & Luce, R. D. (1980). Variability and sequential effects in cross-modality matching of area and loudness. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 277-289.

Berliner, J. E., & Durlach, N. I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America, 53,* 1270-1287.

Berliner, J. E., Durlach, N. I., & Braida, L. D. (1977). Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects. *Journal of the Acoustical Society of America, 61,* 1577-1585.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483-502.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity Perception. XIII. Perceptual anchor model of context-coding. *Journal of the Acoustical Society of America, 76,* 722-731.

DeCarlo, L. T. (1992). Intertrial interval and sequential effects in magnitude scaling. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1080-1088.

DeCarlo, L. T. (1994). A dynamic theory of proportional judgment: Context and judgment of length, heaviness, and roughness. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 372-381.

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General, 119,* 375-396.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372-383.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46,* 373-380.

Garner, W. R. (1954). Context effects and the validity of loudness scales. *Journal of Experimental Psychology, 48,* 218-224.

Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203-216.

Haubensak, G. (1992a). Models for frequency effects in absolute judgment. In G. Borg & G. Neely (Eds.), *Fechner day '92: Proceedings of the eighth annual meeting of the International Society for Psychophysics* (pp. 93-97). Stockholm, Sweden.

Haubensak, G. (1992b). The consistency model: A process model for absolute judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 303-309.

Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception and Psychophysics, 3,* 409-414.

Hu, G. (1997). Why is it difficult to learn absolute judgment tasks? *Perceptual and Motor Skills, 84,* 323-335.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects of the judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92-104.

Kiang, N. Y. S. (1975). Stimulus representation in the discharge pattern of auditory neurons. In E. L. Eagles (Ed.), *The nervous system* (Vol. 3, pp. 81-96). New York: Raven.

Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Research, 60,* 121-133.

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383-395.

Laming, D. R. J. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical and Statistical Psychology, 37,* 152-183.

Laming, D. R. J. (1997). *The measurement of sensation*. London: Oxford University Press.

Lockhead, G. R. (1984). Sequential predictors of choice in psychophysical tasks. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 27-47). Hillsdale, NJ: Erlbaum.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics, 20,* 49-54.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics, 32,* 397-408.

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology, 37,* 136-151.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Mori, S. (1989). A limited-capacity response process in absolute identification. *Perception & Psychophysics, 46,* 167-173.

Mori, S., & Ward, L. M. (1995). Pure feedback effects in absolute identification. *Perception & Psychophysics, 57,* 1065-1079.

Morris, R. B., & Rule, S. J. (1988). Sequential judgment effects in magnitude estimation. *Canadian Journal of Psychology, 42,* 69-77.

Nosofsky, R. M. (1983). Shifts of attention in the identification and discrimination of intensity. *Perception & Psychophysics, 33,* 103-112.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39-57.

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

Petzold, P., & Haubensak, G. (2001). Higher order sequential effects in psychophysical judgments. *Perception & Psychophysics, 63,* 969-978.

Petzold, P., & Haubensak, G. (in press). Local frames of reference: A multiple standards model. In S. Kaernbach (Ed.), *Psychophysics beyond sensation.* Hillsdale, NJ: Erlbaum.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24,* 745-749.

Pollack, I. (1953). The information of elementary auditory displays. II. *Journal of the Acoustical Society of America, 25,* 765-769.

Purks, S. R., Callahan, D. J., Braida, L. D., & Durlach, N. I. (1980). Intensity perception. X. Effect of preceding stimulus on identification performance. *Journal of the Acoustical Society of America, 67,* 634-637.

Schifferstein, H. J. N., & Frijters, J. E. R. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception & Psychophysics, 52,* 243-255.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review, 101,* 357-361.

Staddon, J. E. R., King, M., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception & Performance, 6,* 290-301.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153-181.

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, 28, 3-11.

Stewart, N., & Brown, G. D. A. (2003). *The relationship between similarity-based and difference-based models of perceptual categorization.* Manuscript submitted for publication.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113,* 345-361.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68-111.

Ward, L. M. (1972). Category judgments of loudness in the absence of an experimenter-induced identification function: Sequential effects and power-function fit. *Journal of Experimental Psychology, 94,* 179-184.

Ward, L. M. (1973). Repeated magnitude estimated with a variable standard: Sequential effects and other properties. *Perception & Psychophysics, 13,* 193-200.

Ward, L. M. (1982). Mixed modality psychophysical scaling: Sequential dependencies and other properties. *Perception & Psychophysics, 31,* 53-62.

Ward, L. M. (1985). Mixed-modality psychophysical scaling: Inter- and intramodal sequential dependencies as a function of lag. *Perception & Psychophysics, 38,* 512-522.

Ward, L. M. (1987). Remembrance of sounds past: Memory and psychophysical scaling. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 216-227.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effect and memory in category judgment. *Journal of Experimental Psychology, 84,* 27-34.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73-78.

Weber, D. L., Green, D. M., & Luce, R. D. (1977). Effect of practice and distribution of auditory signals on absolute identification. *Perception & Psychophysics, 22,* 223-231.

Author Note

Neil Stewart, Department of Psychology University of Warwick; Gordon D. A. Brown, Department of Psychology University of Warwick.

We would like to thank Stian Reimers for his comments, and Rosalind Ashworth and James Boot for their help running the experiment. This work was supported by Economic and Social Research Council grant R000239351.

Correspondence concerning this article should be addressed to Neil Stewart, Department of Psychology, University of Warwick, Coventry, CV4 7AL, UK. E-mail: neil.stewart@warwick.ac.uk.

Footnotes

[1]Mathematica code is available for these models from Neil Stewart.

[2]The degrees of freedom do not match in these two ANOVAs because of missing values: Because of the truly random selection of stimuli, not all cells in the design contained data from every participant.

[3]In the original category contrast experiment (Stewart, et al., 2002) the category contrast effect was based on 10 participants each responding to 400 borderline stimuli after extreme stimuli. In the present experiment, which was not designed specifically to detect this effect, the category contrast effect is based upon data from 16 participants each making, on average, only 23.8 critical responses. The reduced power in the current design is probably responsible for the failure to replicate the category contrast effect using Stewart et al.'s (2002) analysis.

An alternative suggestion is that the category contrast effect in the current data might be smaller because the spacing between the stimuli was increased from 1% in Stewart et al.'s (2002) study to 6% in the current study. However, this increase in spacing had only a very slight effect on the proportion of correct responses averaged over all stimuli (84.1% originally verses 85% now). This is in line with similar results in absolute identification that show that, once adjacent stimuli are discriminable, increasing the range has only a very small effect (Braida & Durlach, 1972; Pollack, 1952). We conclude that it is the lack of power rather than the altered stimulus spacing that is responsible for the marginal category contrast effect seen here.

[4]It is not possible to examine the complete three-way interaction between S_n, S_{n-1}, and S_{n-2}, as there are insufficient data. To collect sufficient data, each participant would need to complete approximately 10,000 trials.

[5]There is some evidence that the magnitude of the difference between stimuli might be being used. Performance on the extreme stimuli, Stimuli 1 and 10, is good, no matter what the

preceding stimulus is. The Sign-Only MAC models predict that performance should be better when the preceding stimulus is from the same category (as these stimuli will always be sign-useful) compared to the case when the preceding stimulus is from the opposite category (as these stimuli will never be sign-useful). This prediction is dependent on other recent stimuli also not being sign-useful. As the probability that other stimuli can be recalled rises, this effect is predicted to get smaller. However, if the magnitude of the difference is also available, then the preceding stimulus will either be sign-useful (if it is from the same category) or sufficiently different to promote a swap from the previous feedback to the current response (if it is from the opposite category) (see Figure 4A). If only coarse difference magnitude information is available, as Laming (1997) suggests, this would probably be sufficient. We would like to thank Shuji Mori for this observation.

[6]In addition to this point, there is good neurophysiological evidence that absolute magnitudes are represented in sensory pathways. For example, auditory nerve fibers are tuned to specific frequencies (e.g., Kiang, 1975) and neurons in the auditory cortex are arranged in tonotopic maps (e.g., Abeles & Goldstein, 1970).

[7]An alternative is that these manipulations encourage a more even weighting of recent trials (i.e., S_{n-1}, S_{n-2}, S_{n-3}, ...). On average, this would appear as a reduced effect of trial n-1 and an increased contribution from a long-term frame of reference.

[8]The issue of whether $S_n - S_{n-2}$ is available directly or must be deduced by summing $S_{n-2} - S_{n-1}$ and $S_{n-1} - S_n$ can be determined by examining the variability in responding on trials n-1 and n. The argument is as follows. If $S_n - S_{n-2}$ is the sum of $S_{n-2} - S_{n-1}$ and $S_{n-1} - S_n$ then $\mathrm{Var}(S_n - S_{n-2}) = \mathrm{Var}(S_{n-2} - S_{n-1}) + \mathrm{Var}(S_{n-1} - S_n)$. Thus $\mathrm{Var}(S_n - S_{n-2}) \geq \mathrm{Var}(S_{n-1} - S_n)$. Thus, if information from trial n-2 is used in addition to information from trial n-1, it cannot reduce the variability in responding to S_n if any weighted average of the two sources of information is used (as in Equation 2). This is, of course, not necessarily the case if $S_n - S_{n-2}$ and $S_{n-1} - S_n$ interact in producing R_n (as in the Experiment and Petzold & Haubensak, 2001).

Table 1

A: States of the Availability of Previous Stimuli

Stimulus Availability			
S_{n-1}	S_{n-2}	S_{n-3}	State Probability
0	0	0	$(1 - p_1)(1 - p_2)(1 - p_3)$
0	0	1	$(1 - p_1)(1 - p_2)\, p_3$
0	1	0	$(1 - p_1)\, p_2 (1 - p_3)$
0	1	1	$(1 - p_1)\, p_2\, p_3$
1	0	0	$p_1 (1 - p_2)(1 - p_3)$
1	0	1	$p_1 (1 - p_2)\, p_3$
1	1	0	$p_1\, p_2 (1 - p_3)$
1	1	1	$p_1\, p_2\, p_3$

Note. 0 denotes 'unavailable' and 1 denotes 'available'.

B: An Example for the Category Structure in Figure 1 when $S_n = 3$

Stimulus Availability					
$S_{n-1} = 7$	$S_{n-2} = 5$	$S_{n-3} = 2$	State Probability ($p_1 = .9$, $p_2 = .6$, $p_3 = .3$)	Sign-Useful Stimulus Recalled?	Most Recent Available Stimulus
0	0	0	.028	No	None
0	0	1	.012	No	S_{n-3}
0	1	0	.042	Yes, S_{n-2}	S_{n-2}
0	1	1	.018	Yes, S_{n-2}	S_{n-2}
1	0	0	.252	No	S_{n-1}
1	0	1	.108	No	S_{n-1}
1	1	0	.378	Yes, S_{n-2}	S_{n-1}
1	1	1	.162	Yes, S_{n-2}	S_{n-1}

Table 2

Inferring Long-Range Ordinal Relationships from Consecutive Ordinal Judgments

		Implied Sign of
S_{n-2} - S_{n-1}	S_{n-1} - S_n	S_{n-2} - S_n
+	+	+
-	+	?
+	-	?
-	-	-

Figure Captions

Figure 1. Ten stimuli distributed evenly along a single psychological dimension divided into two categories.

Figure 2. Predictions of the Feedback Repetition MAC Model (b). A: The probability of a correct R_n as a function of S_{n-1} for different S_n. B: The probability of a correct R_n as a function of S_{n-2} for different S_n. C: The probability of a correct R_n as a function of S_{n-1} and S_{n-2} when S_n = 4.

Figure 3. Predictions of the Recalled Stimulus MAC Model (c). A: The probability of a correct r R_n as a function of S_{n-1} for different S_n. B: The probability of a correct R_n as a function of S_{n-2} for different S_n. C: The probability of a correct R_n as a function of S_{n-1} and S_{n-2} when S_n = 4.

Figure 4. Predictions of the Sign and Magnitude MAC Model (d). A: The probability of a correct R_n as a function of S_{n-1} for different S_n. B: The probability of a correct R_n as a function of S_{n-2} for different S_n. C: The probability of a correct R_n as a function of S_{n-1} and S_{n-2} when S_n = 4.

Figure 5. The mean proportion of correct R_n as a function of S_n in the Experiment. Performance has been collapsed across categories (i.e., S_n = 1 represents performance averaged across Stimuli 1 and 10, S_n = 2 represents performance averaged across Stimuli 2 and 9, and so on). Error bars represent the standard errors of the means.

Figure 6. Sequential effects in the Experiment. A: The mean proportion of correct R_n as a function of S_{n-1} for different S_n. B: The mean proportion of correct R_n as a function of S_{n-}2 for different S_n. Performance has been collapsed across categories. Error bars represent the standard errors of the means.

Figure 7. Conditional sequential effects in the Experiment. A: The mean proportion of correct R_n only when the S_{n-1} is sign-useful as a function of S_{n-2} for different S_n. B: The mean proportion of correct R_n only when the S_{n-1} is not sign-useful as a function of S_{n-2} for different

S_n. C: The mean proportion of correct R_n only when the S_{n-2} is sign-useful as a function of S_{n-1} for different S_n. D: The mean proportion of correct R_n only when the S_{n-2} is not sign-useful as a function of S_{n-1} for different S_n.

Figure 1

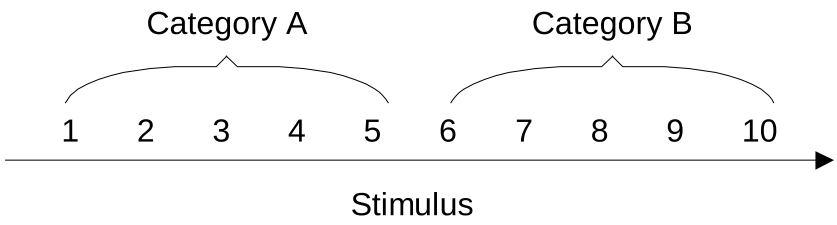


Figure 2

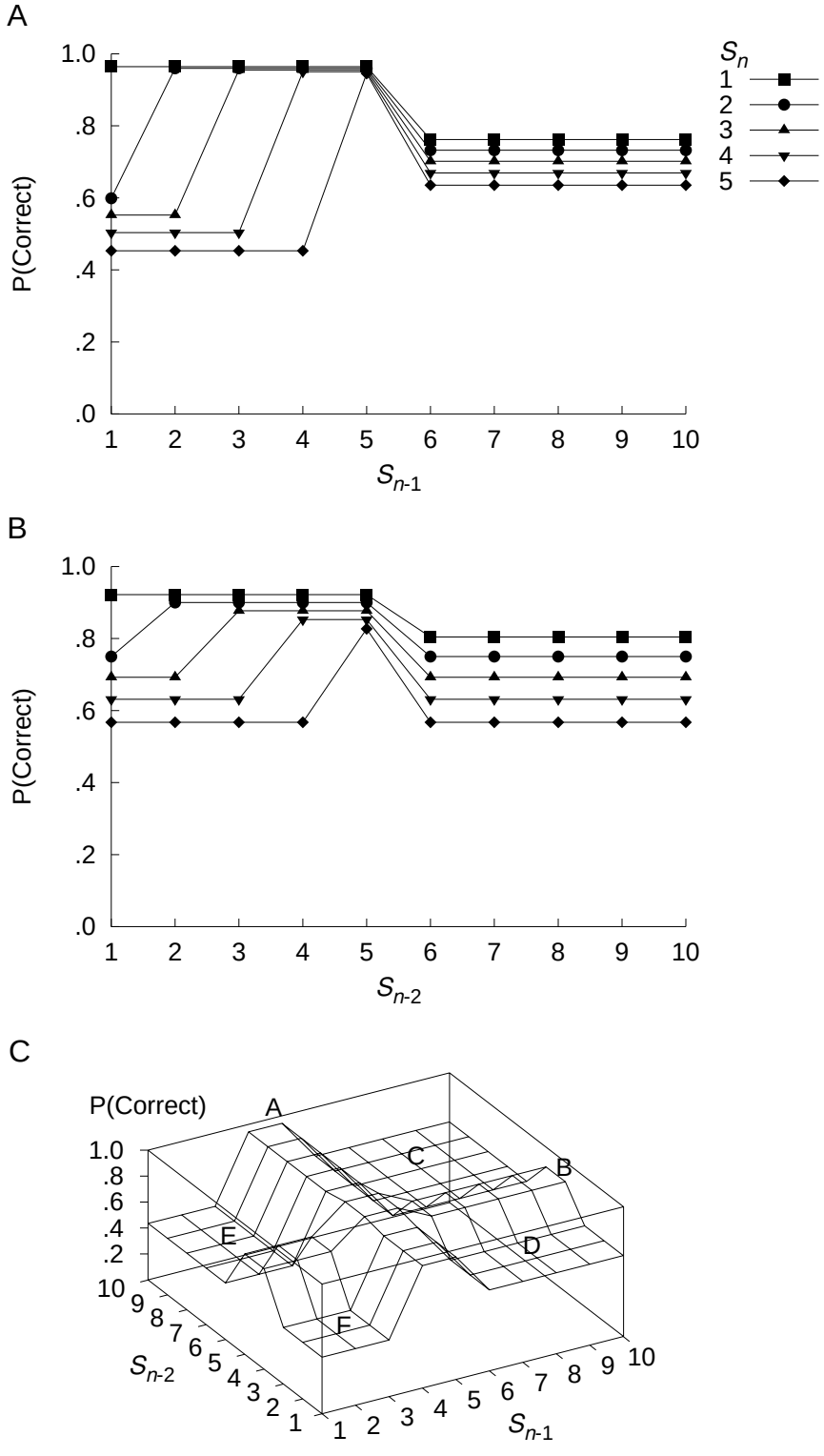

A
P(Correct)
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-1}
S_n
1
2
3
4
5
B
P(Correct)
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-2}
C
P(Correct)
A
B
C
D
E
F
1.0
.8
.6
.4
.2
10 9 8 7 6 5 4 3 2 1
S_{n-2}
1 2 3 4 5 6 7 8 9 10
S_{n-1}

Figure 3

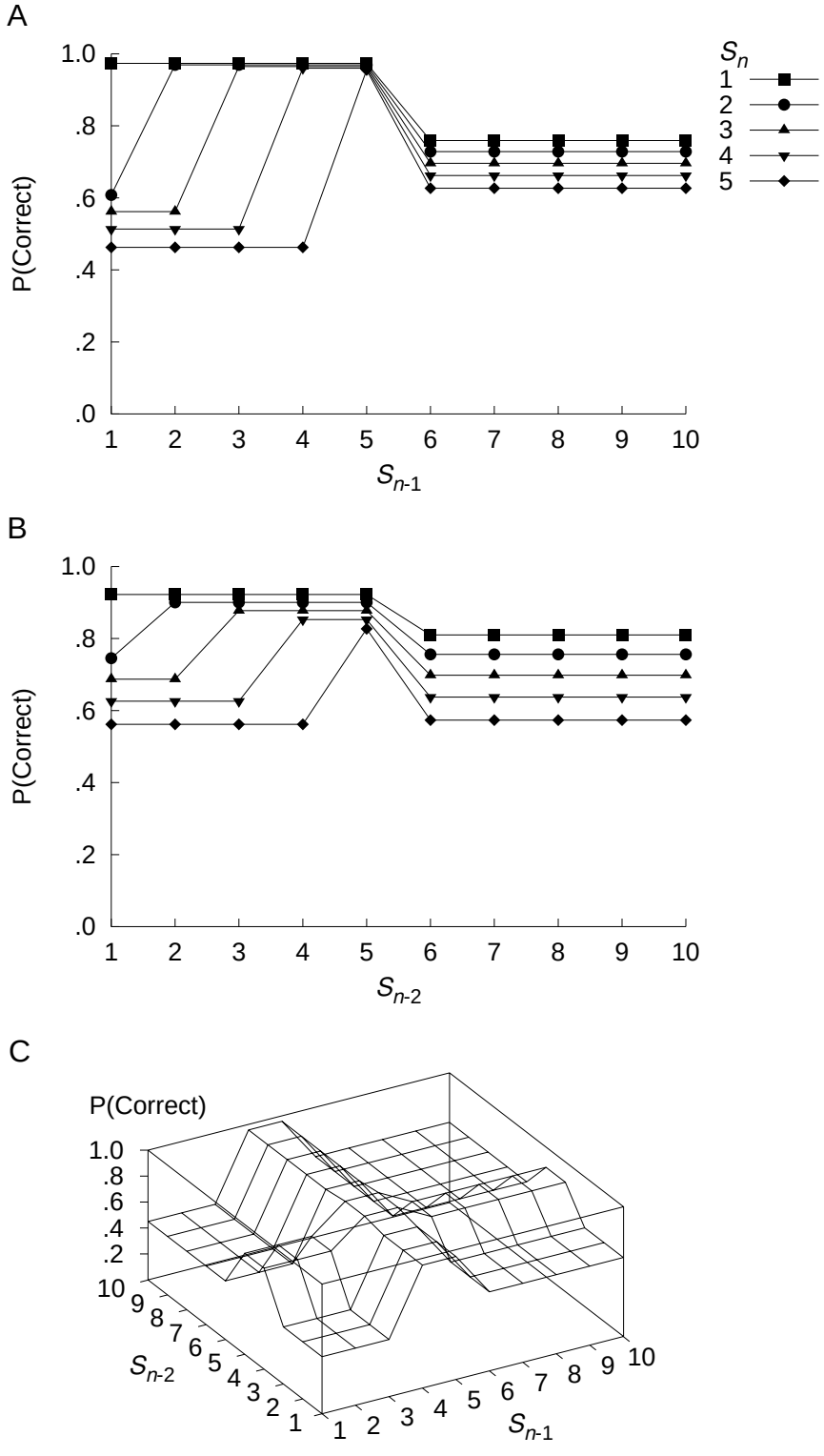

A
1.0
.8
.6
.4
.2
.0
P(Correct)
1 2 3 4 5 6 7 8 9 10
S_{n-1}
S_n
1
2
3
4
5
B
1.0
.8
.6
.4
.2
.0
P(Correct)
1 2 3 4 5 6 7 8 9 10
S_{n-2}
C
P(Correct)
1.0
.8
.6
.4
.2
10 9 8 7 6 5 4 3 2 1
S_{n-2}
1 2 3 4 5 6 7 8 9 10
S_{n-1}

Figure 4

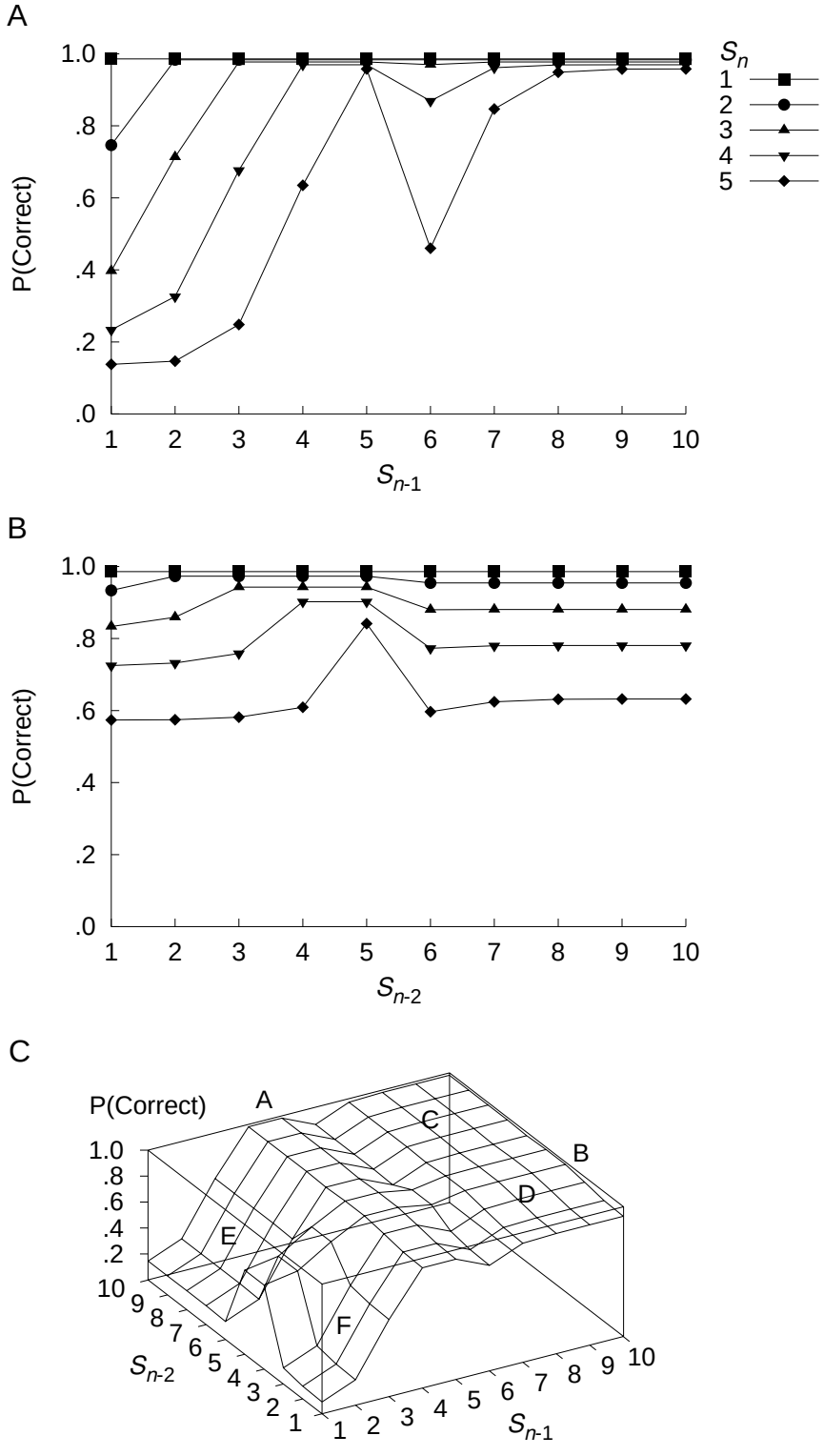

A
P(Correct)
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-1}
S_n
1
2
3
4
5
B
P(Correct)
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-2}
C
P(Correct)
A
B
C
D
E
F
1.0
.8
.6
.4
.2
10 9 8 7 6 5 4 3 2 1
S_{n-2}
1 2 3 4 5 6 7 8 9 10
S_{n-1}

Figure 5

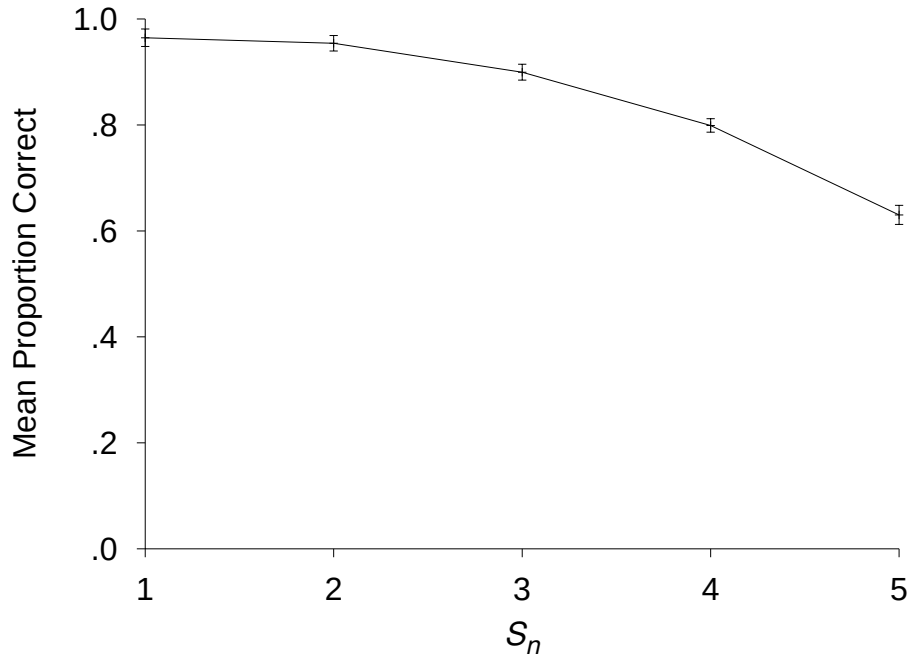


Figure 6

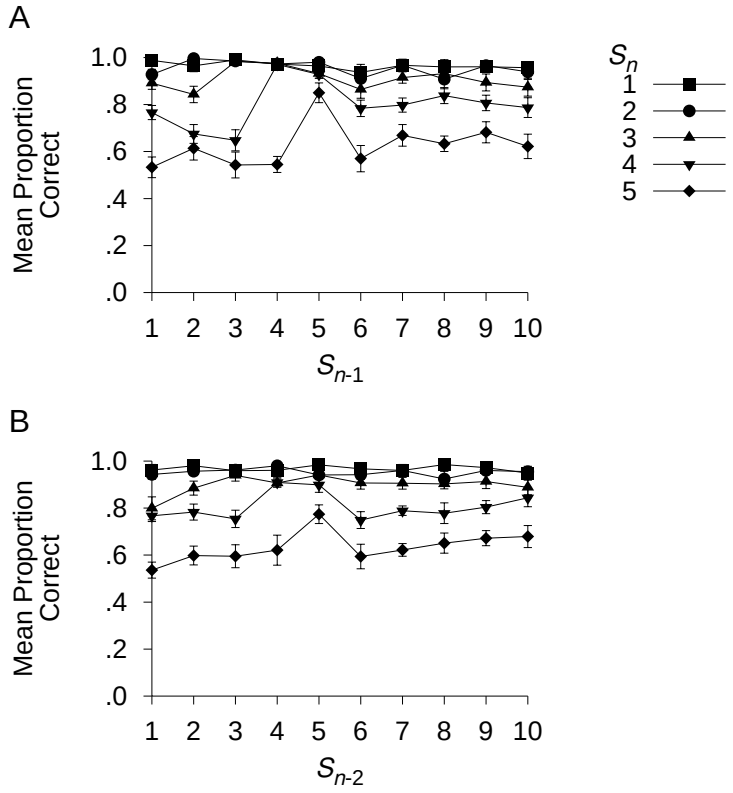
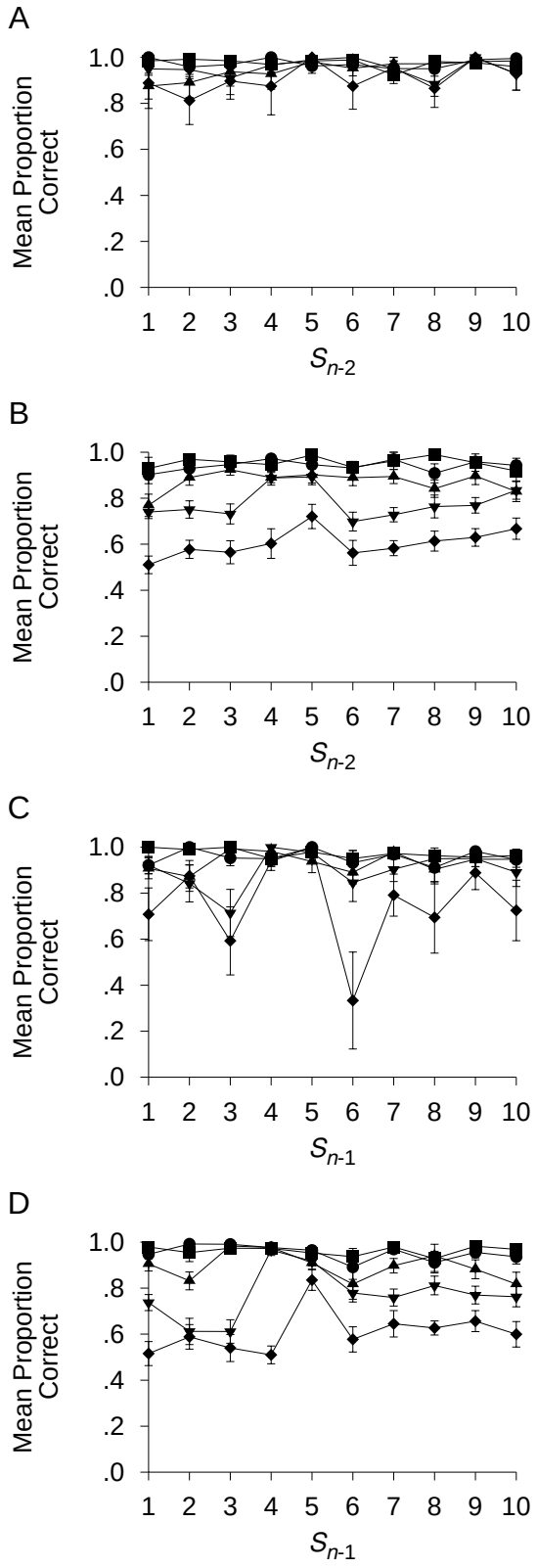

A
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-1}
S_n
1
2
3
4
5
B
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-2}

Figure 7


A
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-2}
B
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-2}
C
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-1}
D
Mean Proportion Correct
1.0
.8
.6
.4
.2
.0
1 2 3 4 5 6 7 8 9 10
S_{n-1}