

# Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks

William J. Matthews and Neil Stewart\*

## Abstract

When participants in psychophysical experiments are asked to estimate or identify stimuli which differ on a single physical dimension, their judgments are influenced by the local experimental context — the item presented and judgment made on the previous trial. It has been suggested that similar sequential effects occur in more naturalistic, real-world judgments. In three experiments we asked participants to judge the prices of a sequence of items. In Experiment 1, judgments were biased towards the previous response (assimilation) but away from the true value of the previous item (contrast), a pattern which matches that found in psychophysical research. In Experiments 2A and 2B, we manipulated the provision of feedback and the expertise of the participants, and found that feedback reduced the effect of the previous judgment and shifted the effect of the previous item's true price from contrast to assimilation. Finally, in all three experiments we found that judgments were biased towards the centre of the range, a phenomenon known as the "regression effect" in psychophysics. These results suggest that the most recently-presented item is a point of reference for the current judgment. The findings inform our understanding of the judgment process, constrain the explanations for local context effects put forward by psychophysicists, and carry practical importance for real-world situations in which contextual bias may degrade the accuracy of judgments.

Keywords: sequential effects; magnitude estimation; price; anchoring

## 1 Introduction

When people make basic perceptual judgments — about the brightness of a light or the loudness of a tone, for example — their responses are greatly influenced by the context in which the stimuli are presented: A square of a given size is regarded as large when most of the squares in the experiment are of smaller size, but regarded as small when most of the squares are larger (e.g., Parducci, 1965). It has long been known that this insight from psychophysics has relevance to more complex, "real world" situations. Thus, the perceived severity of a moral transgression depends upon the ensemble of scenarios presented for judgment, even when participants are explicitly instructed to ignore this context (Parducci, 1968).

Psychophysical studies have therefore demonstrated the importance of what may be termed the global experimental context — the set of stimuli employed — in determining perceptual judgments, and this insight has proven useful in more complex and naturalistic judgment tasks. Yet psychophysical judgments are also influenced by lo-

cal context — by the stimuli presented on the past few trials — and a number of authors have provided evidence that the same principal applies to more complex, non-perceptual decisions too (e.g., Beckstead, 2008; Vlaev & Chater, 2007). The current article develops this idea by studying sequential effects in a complex judgment task using the analytical tools and experimental manipulations employed in psychophysical research.

The dependency of perceptual judgments on the events of the last few trials has been extensively researched by psychophysicists. The typical approach is to employ a regression model in which the current judgment,  $J_n$ , is the dependent variable and the current stimulus and stimuli and/or responses from trials earlier in the sequence are predictors. In particular, Jesteadt, Luce, and Green (1977) advocated the use of the following regression model:

$$J_n = \gamma + \alpha_0 P_n + \alpha_1 P_{n-1} + \beta_1 J_{n-1} + \epsilon \quad (1)$$

where  $P_n$  is the value of the stimulus presented on the current trial,  $P_{n-1}$  is the value of the stimulus on the previous trial, and  $J_{n-1}$  is the value of the judgment made on the previous trial. Equation 1 has been applied to data from a large number of psychophysical experiments. In these experiments, the participant is presented with a sequence of stimuli which differ in one physical attribute,

\*This work was supported by Economic and Social Research Council grant RES-000-23-1372. We thank Malik Refaat for help with data collection. Address correspondence to: William Matthews, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom. Email: w.j.matthews@warwick.ac.uk. Data for this study are available at <http://journal.sjdm.org/81104/data.html>.

such as tones which differ in loudness, and asked to form some judgment of that attribute; the precise nature of the judgment depends on the psychophysical task. In magnitude estimation experiments, the participant is asked to assign a number which indicates his or her subjective impression of the loudness of each tone, either with respect to an explicit standard (e.g., Reynolds & Stevens, 1960) or on an absolute scale (e.g., Ward, 1987). In cross-modality matching experiments, the participant is asked to adjust the magnitude of one dimension, such as loudness, so that it matches a magnitude on another dimension, such as brightness (e.g., Ward, 1979). In category judgment experiments, the participant is asked to put each stimulus into one of several categories (such as “very quiet,” “quiet,” “medium,” “loud,” “very loud”; e.g., Petzold & Haubensak, 2001). In absolute identification experiments, each stimulus is given a unique label — for example, the stimuli are numbered 1–10 — and the participant is asked to name the stimulus presented on each trial (e.g., Garner, 1953).

Equation 1 has been used to study sequential effects in all of these paradigms. The details of the results depend somewhat on the experimental task but the general pattern is robust: The response on the current trial is biased towards the judgment made on the previous trial but away from the stimulus presented on that trial (e.g., DeCarlo & Cross, 1990; Jesteadt et al., 1977; Matthews & Stewart, in press; Mori, 1998; Mori & Ward, 1995; Ward, 1979; 1987). That is, there is assimilation to the immediately preceding response but contrast to the immediately preceding stimulus.

These sequential effects have been given various interpretations, many of which assume that there is some kind of perceptual interference from the previous stimulus and that the previous item and the judgment assigned to it serve as a point of reference when evaluating the current stimulus (e.g., DeCarlo & Cross, 1990). It is argued that even when participants are asked to judge stimuli with respect to long-term referents, they use the most recently experienced events as a framework for judgment (Holland & Lockhead, 1968; Laming, 1984; Stewart, G. D. A. Brown, & Chater, 2005).

Many real-world tasks have a structural similarity to magnitude estimation or category judgment, in that people estimate or classify a sequence of stimuli. However, the stimuli are very different. The tones, lights and lines used in psychophysical investigations are very simple and notoriously difficult to store in long-term memory; indeed, it is frequently asserted that our capacity for processing such stimuli is limited to about 7 items (Miller, 1956), in contrast to our capacity to recognise and identify many thousands of complex objects (e.g., Matthews, Benjamin, & Osborne, 2007). The labile mental representations of psychophysical stimuli may be responsible

for the observed sequential effects; the inability to form accurate long-term representations may push people towards the use of recent items as a frame of reference. When people make real-world judgments about complex items, and when the judgments are of a type with which they are very familiar, it may be that the sequential effects are eliminated as people use only long-term referents and stable internal scales of judgment. Indeed, many of the models of sequential effects in psychophysical judgment make assumptions which explicitly concern perceptual tasks and which are not readily extended to other situations (e.g., S. D. Brown, Marley, Donkin, & Heathcote, 2008).

The current article asks whether the pattern of sequential dependencies seen in psychophysical tasks extends to situations in which people make judgments about non-physical dimensions of complex, real-world objects. In three experiments we asked participants to judge the prices of various items. We chose this task because it corresponds reasonably well to an important aspect of our economic lives; we are routinely exposed to sequences of products and, implicitly or explicitly, assess their probable cost. This task was also attractive because it allowed us to use rich, complex stimuli and to require judgments about a property that does not correspond to a simple physical aspect of the item presented for judgment.

In Experiment 1, we ask whether judgments of price exhibit sequential dependencies of the type seen in psychophysical experiments. In Experiments 2A and 2B we extend the results of Experiment 1 to new stimuli and a modified procedure, and ask whether experimental manipulations known to influence sequential dependencies in psychophysical tasks exert the same effects on judgments of price.

## 2 Experiment 1

In Experiment 1 participants judged the prices of chairs. We chose chairs as the to-be-judged items because we felt their prices should be relatively obvious from their appearance (unlike, say, electronic goods, which may have many hidden features). We presented pictures of chairs taken from the website of a popular retailer (Ikea). One can browse this website and purchase chairs based entirely upon their photographs.

### 2.1 Method

#### 2.1.1 Participants

Twenty five staff and students from the University of Warwick took part; each was paid £3. All participants had been resident in the UK for at least the past 3 years.

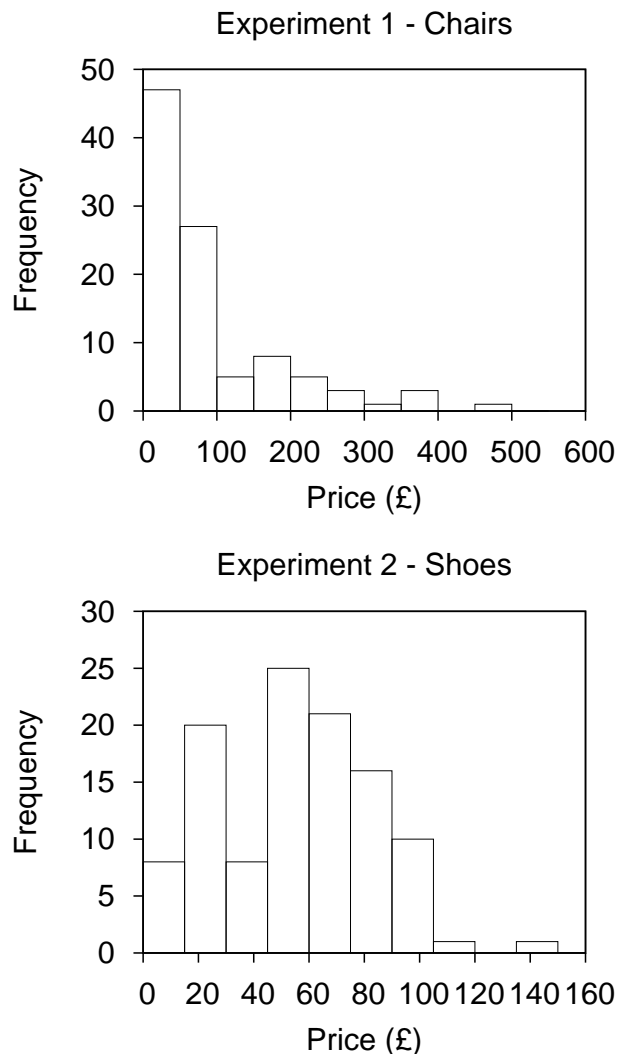


Figure 1: Distribution of prices of stimuli in Experiment 1 (left) and Experiments 2A and 2B (right).

### 2.1.2 Stimuli

The stimuli were 100 pictures of chairs available from the Ikea furniture store. The prices of the chairs ranged from £6.49 to £489 ( $M = 95.49$ ,  $SD = 98.86$ ); the distribution of prices is shown in Figure 1. Each picture measured 500x500 pixels and was presented on a 19" TFT monitor with resolution 1280x1024 pixels, viewed from approximately 50cm.

### 2.1.3 Procedure

Participants were tested in individual testing cubicles. On each trial, participants were shown one of the chairs. Underneath the photo was a box in which the participant typed his or her estimate of the price; participants were asked to enter the price in pounds and were free to use

the decimal point if they wished. Participants were free to edit their responses (e.g., by deleting the number) and entered their judgment by pressing Enter or Return. The screen then went blank for 500 ms before the presentation of the next chair. Participants were told not to worry if they were uncertain and just to enter their best estimate of each item's price. Each participant judged each item once, giving 100 trials per participant; the order of presentation was randomized for each person.

## 2.2 Results

In this experiment, and in those which follow, there were a small number of missing/nonsensical responses (e.g., "£0") and cases where, after completing the experiment, the participant reported having mis-typed a particular judgment. Such responses were excluded from the analysis, as were a handful of extremely large responses (more than five inter-quartile ranges above the upper quartile for that participant) which were assumed to have been entered in error. In the current experiment a total of 0.12% of responses were excluded.

### 2.2.1 Relationship between true price and judged price

We began by plotting the relationship between judged price and true price separately for each participant. The top panel of Figure 2 shows the results averaged over participants; each point represents the average judgment for a given product.

Judged price increases with true price. However, the distribution of true prices is highly skewed; there are a large number of low-priced products and fewer expensive items (see e.g., Stewart & Simpson, 2007, for other examples of this in price data), giving very high leverage to a relatively small number of expensive items. In addition, there is evidence of a curvilinear relationship between price and judgment, with the curve becoming flatter at higher prices (this pattern was more obvious in individual participant data). We therefore applied a logarithmic transformation to both the true price and judgment values. The results, averaged over participants, are shown in the bottom panel of Figure 2. The transformation brought the data into better agreement with the assumptions of linear regression, so for all subsequent analyses we used the log-transformed data.

To investigate the relationship between price and judgment, we regressed judged prices on true prices. The regression line for the participant-averaged data is shown in the bottom of Figure 2, along with a line of zero intercept and slope of one that indicates the results expected if judgments were perfectly accurate — the "veridical" line. The regression line is swivelled with respect to the veridi-

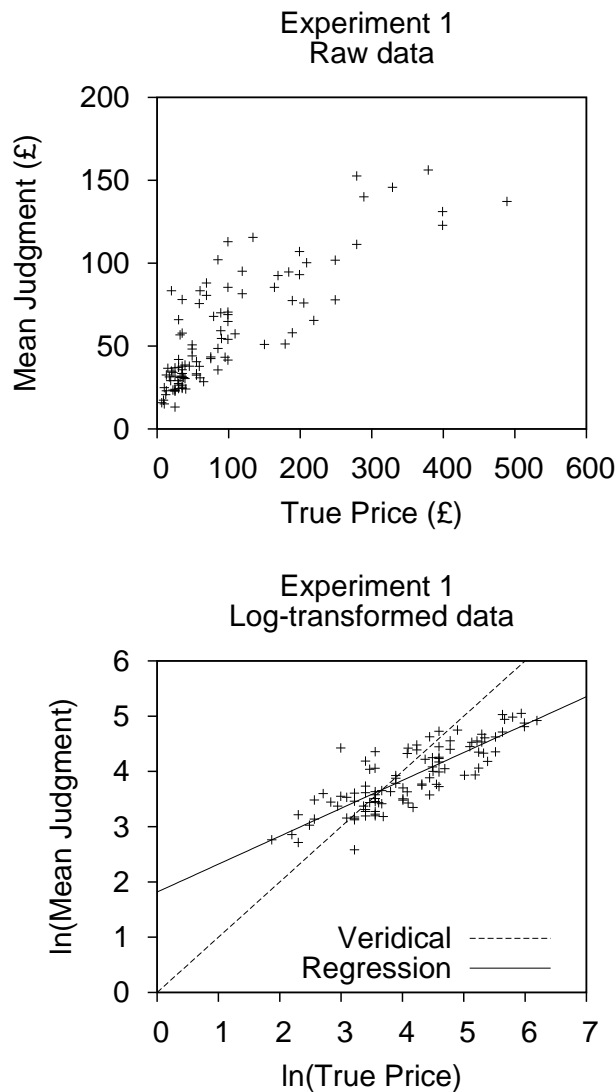


Figure 2: Relationship between true price and judged price in Experiment 1. The plots show the mean judgment for each product. The top panel shows the untransformed data; the bottom panel shows the results after logarithmically transforming both the true prices and the judgments. The solid line in the bottom panel is the regression line, indicating the empirical relationship between true price and judgment. The dashed line shows the results expected if judgments were perfectly accurate.

cal line; on average, participants overestimated the price of cheap items and underestimated the price of expensive items.

We applied the same approach to the data from each individual participant. There was a significant positive relationship between true price and judged price for every participant but, as seen in the averaged data, the regression lines were rotated towards the horizontal. We return

to this point later.

### 2.2.2 Sequential Effects

To examine sequential effects we applied Equation 1; that is, we regressed the judgment on the  $n$ th trial,  $J_n$ , on the true price of the  $n$ th item,  $P_n$ , the true price of the previous item,  $P_{n-1}$ , and the judgment made for the previous item,  $J_{n-1}$  (in all cases using log-transformed variables). We fit the data from each participant separately. The results are shown in Table 1. The leftmost columns of Table 1 contain the unstandardized regression coefficients along with associated significance codes. Some researchers (e.g., Beckstead, 2007) have pointed out that the significance of a regression coefficient depends critically upon the number of experimental trials, and that researchers should consider effect sizes when deciding the importance of a particular judgment cue. The rightmost column of Table 1 therefore contains the standardized coefficients (the significance codes for these are, of course, the same as for the unstandardized values, and are not listed in Table 1).

For every participant there is a significant positive relationship between actual price and judged price. There is also evidence of sequential effects, particularly of the preceding judgment. For 23 of the 25 participants the coefficient for  $J_{n-1}$  is positive, indicating assimilation to the previous judgment; for 9 of these 23 participants the effect is significant. Neither of the participants with negative  $J_{n-1}$  coefficients show a significant effect. The effects of the preceding item's actual price,  $P_{n-1}$ , are less consistent; 19 of the 25 participants have a negative coefficient, two of which are significant. None of the participants with a positive coefficient show a significant effect.

There is some multicollinearity in the regression model because the  $J_{n-1}$  and  $P_{n-1}$  predictors are correlated. Multicollinearity increases the standard errors of the regression coefficients (although the coefficients remain unbiased estimators), reducing the likelihood that a particular coefficient will be significant. We examined the variance inflation factor (VIF) to assess the severity of the multicollinearity. We calculated the VIFs for the regression analyses from each participant in each experiment (a total of 81 regressions). Each regression had three independent variables, giving a total of 243 VIF values. The results indicated that the multicollinearity was not severe; only 2 of the VIFs were more than 3.0 (and then only slightly, 3.01 and 3.03). Some textbooks have suggested that VIFs less than 10.0 indicate acceptable multicollinearity (e.g., Neter, Kutner, Nachtsheim, & Wasserman, 1996). Nonetheless, some caution is needed when interpreting the significance of results for individual participants, and in order to establish the overall pattern, we used a one-sample  $t$ -test to

Table 1: Regression coefficients for Experiment 1.

Participant	Unstandardized				Standardized			$R^2$
	int	$\log P_n$	$\log P_{n-1}$	$\log J_{n-1}$	$\log P_n$	$\log P_{n-1}$	$\log J_{n-1}$	
1	0.676	0.653 <sup>a</sup>	-0.035	0.137	0.736	-0.039	0.137	0.540
2	2.131 <sup>a</sup>	0.444 <sup>a</sup>	-0.188 <sup>b</sup>	0.200 <sup>d</sup>	0.764	-0.324	0.203	0.626
3	0.928 <sup>c</sup>	0.505 <sup>a</sup>	-0.041	0.235 <sup>d</sup>	0.713	-0.057	0.240	0.521
4	0.582 <sup>d</sup>	0.462 <sup>a</sup>	-0.026	0.221 <sup>d</sup>	0.692	-0.039	0.224	0.524
5	0.864 <sup>c</sup>	0.432 <sup>a</sup>	-0.105	0.338 <sup>b</sup>	0.704	-0.172	0.341	0.561
6	0.977 <sup>c</sup>	0.636 <sup>a</sup>	-0.066	0.118	0.758	-0.079	0.118	0.566
7	1.807 <sup>a</sup>	0.575 <sup>a</sup>	0.074	0.037	0.735	0.094	0.037	0.557
8	1.408 <sup>a</sup>	0.452 <sup>a</sup>	-0.037	0.032	0.760	-0.063	0.032	0.587
9	-0.484	0.732 <sup>a</sup>	0.051	0.161	0.796	0.055	0.161	0.651
10	0.448	0.652 <sup>a</sup>	-0.040	0.044	0.790	-0.048	0.043	0.632
11	1.219 <sup>c</sup>	0.435 <sup>a</sup>	0.129	0.121	0.541	0.163	0.122	0.363
12	1.950 <sup>a</sup>	0.351 <sup>a</sup>	-0.032	0.175	0.605	-0.055	0.176	0.373
13	0.725 <sup>d</sup>	0.546 <sup>a</sup>	-0.017	0.107	0.677	-0.020	0.109	0.480
14	2.979 <sup>a</sup>	0.311 <sup>a</sup>	0.080	-0.168	0.634	0.163	-0.168	0.409
15	2.187 <sup>a</sup>	0.563 <sup>a</sup>	-0.055	0.032	0.668	-0.065	0.032	0.453
16	1.032 <sup>c</sup>	0.455 <sup>a</sup>	-0.174 <sup>c</sup>	0.475 <sup>a</sup>	0.685	-0.261	0.478	0.627
17	0.599	0.542 <sup>a</sup>	-0.111	0.306 <sup>c</sup>	0.704	-0.143	0.308	0.573
18	1.768 <sup>a</sup>	0.608 <sup>a</sup>	-0.046	0.088	0.666	-0.050	0.094	0.454
19	-0.017	0.694 <sup>a</sup>	-0.023	0.363 <sup>a</sup>	0.819	-0.027	0.369	0.677
20	0.183	0.491 <sup>a</sup>	0.031	0.319 <sup>b</sup>	0.615	0.039	0.321	0.512
21	2.287 <sup>a</sup>	0.207 <sup>a</sup>	-0.040	0.147	0.587	-0.112	0.155	0.351
22	1.404 <sup>a</sup>	0.354 <sup>a</sup>	-0.102	0.342 <sup>b</sup>	0.668	-0.194	0.346	0.528
23	1.066 <sup>c</sup>	0.585 <sup>a</sup>	0.009	-0.061	0.717	0.011	-0.061	0.532
24	0.751 <sup>d</sup>	0.604 <sup>a</sup>	-0.052	0.138	0.782	-0.068	0.138	0.607
25	0.741 <sup>d</sup>	0.521 <sup>a</sup>	-0.025	0.179	0.722	-0.035	0.185	0.506
<i>M</i>	1.128	0.512	-0.034	0.163	0.702	-0.053	0.166	0.528
<i>SD</i>	0.797	0.126	0.072	0.143	0.069	0.113	0.145	0.090

<sup>a</sup>  $p < .0001$ , <sup>b</sup>  $p < .001$ , <sup>c</sup>  $p < .01$ , <sup>d</sup>  $p < .05$

Table 2: Mean  $R^2$  changes for each coefficient.

	$P_n$	$P_{n-1}$	$J_{n-1}$	$\frac{P_{n-1}+}{J_{n-1}}$
Experiment 1	0.491	0.013	0.024	0.037
Experiment 2A (No feedback)	0.416	0.011	0.028	0.039
Experiment 2B (Feedback)	0.493	0.026	0.010	0.036

Note: predictors were log-transformed in Experiment 1.

see whether the mean of the (unstandardized) regression coefficients collected from the 25 subjects reliably differed from zero (Lorch & Myers, 1990; see e.g., Ward, 1985; 1987; 1990). For all three predictors ( $P_n$ ,  $P_{n-1}$ , and  $J_{n-1}$ ) the mean coefficients were significantly different from zero. As one would expect, the  $P_n$  coefficient was positive,  $t(24) = 20.4$ ,  $p < .001$ . The  $P_{n-1}$  coefficient was negative,  $t(24) = 2.33$ ,  $p = .029$ , indicating contrast to the true price of the preceding item. Lastly, the  $J_{n-1}$  coefficient was positive,  $t(24) = 5.70$ ,  $p < .001$ , indicating that the current judgment assimilates towards the previous judgment.

We conducted hierarchical regression, entering the predictors in the order  $P_n$ ,  $P_{n-1}$ ,  $J_{n-1}$ , to establish the  $R^2$  increases for each (see Mori & Ward, 1995; Ward, 1979; 1987, for a similar approach). The mean  $R^2$  values (averaged over participants) are shown in Table 2. The proportion of variance attributable to events from the previous trial is not huge, but is comparable to the results from some psychophysical studies. For example, the “high information” condition of Ward’s (1979) magnitude estimation experiment produced a total  $R^2$  change for the events of the preceding trial of .019.

### 2.2.3 Second order effects

Studies of perceptual judgment have found that the correlation between successive judgments depends upon the size of the difference between successive stimuli. When  $P_n$  and  $P_{n-1}$  are similar,  $J_n$  and  $J_{n-1}$  are highly correlated; as  $P_n$  and  $P_{n-1}$  move further apart, the correlation between responses drops away to zero (e.g., Baird, Green, & Luce, 1980; Jesteadt et al., 1977; Ward, 1979; 1982; 1985). Looking for such second order dependencies in our data is problematic because of the small number of trials. Nonetheless, we conducted an exploratory analysis.

The approach we took was based upon that used by Jesteadt et al. (1977; see also Ward, 1979) in studies of magnitude estimation, where the relationship between

stimuli and responses is described by a power law,  $J_n = kP_n^\kappa$ . Jesteadt et al. developed the following approach to second order sequential dependencies. First, regress  $\log J_n$  on  $\log P_n$  separately for each participant to obtain values of  $k$  and  $\kappa$ . Second, normalize each response by dividing it by the value expected on the basis of the stimulus and the overall power function parameters. Third, group the data according to the difference between  $\log P_n$  and  $\log P_{n-1}$  (the jump size). Finally, for each jump size, compute the relationship between the current (normalized) response and the previous one, according to the equation:

$$\log \frac{J_n}{kP_n^\kappa} = \delta + \beta \log \frac{J_{n-1}}{kP_{n-1}^\kappa} + \epsilon \quad (2)$$

We employed the same approach. (Recall that the skewed distribution of prices in this experiment meant that we used log-transformed variables, so a straightforward implementation of Jesteadt et al.’s (1977) approach is appropriate.) We had to aggregate across jump sizes to obtain a useable number of trials in each condition; we placed the values of  $\log P_n - \log P_{n-1}$  into five bins such that, across the whole experiment, an approximately equal number of observations fell into each bin. For each bin size we calculated, separately for each participant, the correlation between  $\log(J_n/kP_n^\kappa)$  and  $\log(J_{n-1}/kP_{n-1}^\kappa)$ . Figure 3 shows the mean correlation coefficients. As one would expect with so few data points, the correlation coefficients are very noisy, and a one-way within subjects ANOVA indicated no significant effect of jump size,  $F(4, 96) < 1$ . However, inspection of the figure suggests slight evidence for the inverted v-shaped pattern seen in psychophysical studies.

### 2.2.4 Depth of sequential effects

Finally, we tried fitting a regression model which included the events from two trials back as predictors; there was no evidence for a consistent effect of  $P_{n-2}$  or  $J_{n-2}$ . The coefficients were significant for only a handful of participants (three showed a significant positive effect of  $J_{n-2}$ ; none showed a significant effect of  $P_{n-2}$ ) and one sample  $t$ -tests on the mean coefficients revealed no significant effect of  $P_{n-2}$  ( $M = 0.006$ ,  $SD = 0.068$ ,  $t(24) < 1$ ) or  $J_{n-2}$  ( $M = 0.040$ ,  $SD = 0.112$ ,  $t(24) = 1.77$ ,  $p = .089$ ). Although these results hint that there might be an effect of  $J_{n-2}$  on the current judgment, this was not borne out in the subsequent experiments.

## 2.3 Discussion

The results of Experiment 1 may be summarized as follows. Firstly, participants’ judgments were correlated with the true prices of the items, but this correlation was

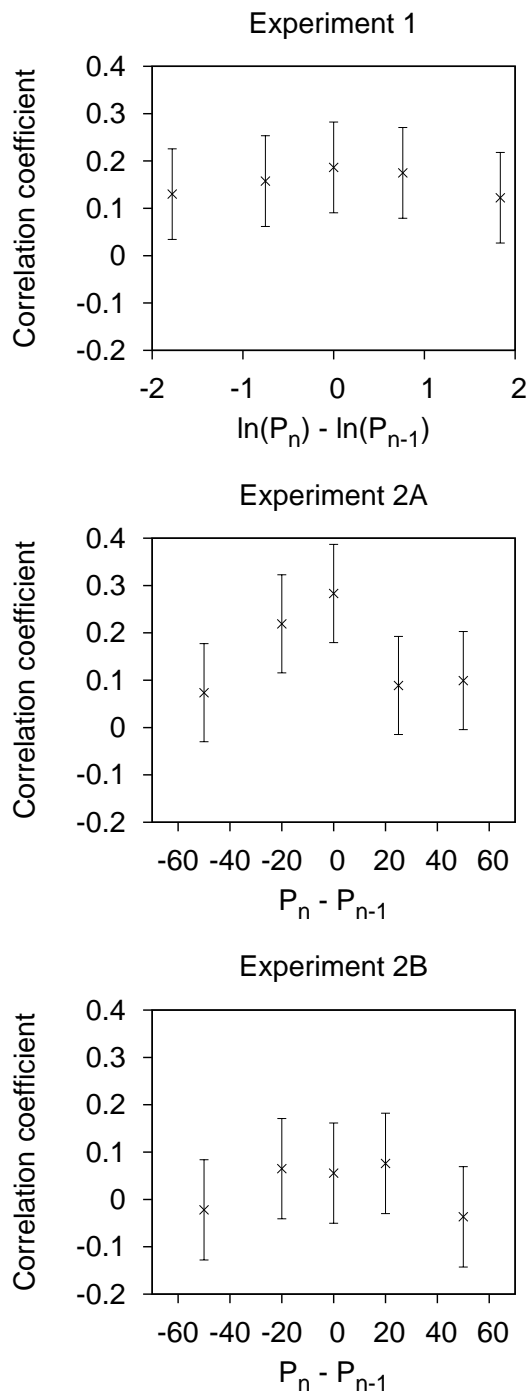


Figure 3: Effect of jump size on correlation between successive responses. Error bars show 95% confidence intervals calculated for a within-subject design (Masson & Loftus, 2003). Note that in Experiment 1 (top panel), we used log-transformed price and judgment values; in Experiments 2A and 2B we used raw values.

far from perfect. Secondly, the regression line relating judgment to true price was rotated towards the horizon-

tal; participant's judgments were biased towards the centre of the range. Thirdly, and most importantly, there were sequential effects; the judgment on the  $n$ th trial depended on the events of the previous trial. Specifically, the current judgment was biased away from the true price of the previous item but towards the judged price of that item; this latter effect was particularly pronounced. These results mimic those found in psychophysical judgments (e.g., Jesteadt et al., 1977; Mori, 1998; Ward, 1987) and suggest that, just as for judgments about the physical aspects of very simple stimuli, the immediate experimental context provides an important influence on complex judgments about rich, multidimensional objects.

Experiment 1 has several limitations. The principal problem is that the observed sequential dependencies are open to a number of interpretations. It may be that the assimilation to the previous judgment indicates the use of that judgment as a point of reference (e.g., Laming, 1995). However, response assimilation might also appear if participants simply have a tendency to repeat the previous response or a reluctance to move very far along the judgment scale, as may occur if they are not fully engaged with the task. In addition, it is unclear whether the results of Experiment 1 are specific to the stimuli and procedure employed in this study.

We therefore conducted two more experiments using a different class of product (women's footwear) and a slightly different experimental procedure. The two new experiments differed in whether or not the participants were told the correct value of each product after they had entered their judgment. In Experiment 2A, no such trial-by-trial feedback was provided; in Experiment 2B, feedback was provided after every judgment. If the sequential effects found in Experiment 1 arise because the previous trial serves as a point of reference, the provision of feedback should exert a pronounced influence on the form of these effects. In psychophysical tasks, feedback reduces the dependency on the previous judgment and increases the dependency on the preceding stimulus (e.g., Mori & Ward, 1995; Ward & Lockhead 1971), presumably because participants now use the true value of the previous item, rather than their own judgment of it, as a point of reference (e.g., Stewart et al., 2005). If, on the other hand, the response assimilation observed in Experiment 1 results from some non-specific effect, such as a tendency to repeat responses, the provision of feedback should make little difference.

As an additional manipulation, we asked whether expertise influences the form or magnitude of the sequential dependencies. It seems plausible that participants who know more about the product being judged will be less reliant on short-term comparisons and less likely to show sequential effects; improving information about the stimulus reduces the sequential dependencies in mag-

nitude estimation and perceptual identification (Mori, 1998; Ward, 1979). Since the products to be judged were items of women's footwear, we examined the issue of expertise by comparing the results from male and female participants.

### 3 Experiment 2A

Whereas in Experiment 1 the item to be judged stayed on-screen until the participant entered a judgment, Experiment 2A presented each item for a fixed time (3s) and then provided a fixed window for the participant to enter his or her judgment (another 3s). The time between successive items was therefore fixed, which is potentially important in judgment tasks (e.g., DeCarlo, 1992; Matthews & Stewart, in press).

#### 3.1 Method

##### 3.1.1 Participants

Twenty eight participants took part, 14 males aged 19-26 years ( $M = 21.0$ ,  $SD = 1.9$ ) and 14 females aged 19-22 ( $M = 20.1$ ,  $SD = 0.9$ ). All had been resident in the United Kingdom for at least the past 3 years.

##### 3.1.2 Stimuli

The stimuli were pictures of 110 items of women's footwear available from a popular high-street chain (Topshop). The prices of the items ranged from £6 to £140 ( $M = 58.33$ ,  $SD = 27.27$ ); the distribution of prices is shown in Figure 1. The pictures were sampled from the Topshop website. All pictures measured 500x500 pixels and showed a single item of footwear on a white background. The stimuli were shown on a 19" TFT monitor with a resolution of 1280x1024 pixels viewed from approximately 50cm.

##### 3.1.3 Design and Procedure

On each trial the participant was shown one picture and asked to judge the price of the item. Each item was shown for 3s and followed by a 3s window during which the participant typed his or her judgment. At the end of the response window the screen went blank for 1.5s before presentation of the next item. Participants completed eight practice trials followed by three blocks of 34 test trials, and were allowed to take a short break between blocks. As the true prices of the items were always integers, participants were instructed to enter their judgments to the nearest pound. The order in which the 110 items were presented for judgment was randomized for each participant.

### 3.2 Results and Discussion

The 8 practice trials were excluded from analysis. Participants failed to enter a response within the 3s window on a small minority of trials (2.2%). A handful of additional responses (0.28% of the total test trials) were excluded for the reasons described in Experiment 1.

#### 3.2.1 Relationship between actual price and judged price

As can be seen in Figure 1, the distribution of prices was much less skewed than in Experiment 1. (This more even distribution was fortuitous; the items were sampled at random from the Topshop website.) The top panel of Figure 4 plots the mean judgment for each item against the item's true price, and indicates a linear relationship between price and judgment, with errors which do not systematically vary with price. The results from individual participants showed the same patterns, and we therefore used untransformed prices and judgments in the regression analyses for this experiment.

We regressed judged price on true price for each participant. The coefficients were positive for all participants and significant for all but one. The participant-averaged data in Figure 4 suggest over-estimation of cheap products and under-estimation of expensive ones. The same pattern appeared in individual participants' data.

#### 3.2.2 Sequential Effects

As before, we examined sequential effects by fitting the regression model described in Equation 1. (The first trial of each block was omitted from this analysis because the time since the presentation of the most recent item depended on how long a break the participant took between blocks.) The regression coefficients for each participant are shown in Table 3.

For every participant but one there is a significant positive relationship between true price and judged price. For the  $P_{n-1}$  term, 17 coefficients are negative (3 significant) and 11 are positive (1 significant). For the  $J_{n-1}$  term, 26 are positive (12 significant) and 2 are negative (neither significant). One sample  $t$ -tests on the coefficients confirmed the impression given by the results from individual participants: There was a significant positive dependence of judgment on actual price,  $t(27) = 16.1$ ,  $p < .001$ , a significant negative dependence on the previous item's price,  $t(27) = 2.57$ ,  $p = .016$ , and a significant positive dependence on the previous judgment,  $t(27) = 6.51$ ,  $p < .001$ .

Unlike Experiment 1, the time between trials was constant in this experiment (with the exception of the self-paced breaks between blocks), so we can use trial number as a measure of time and ask whether the observed



Table 3: Regression coefficients for Experiment 2A (No feedback).

Participant	Unstandardized				Standardized			$R^2$
	int	$P_n$	$P_{n-1}$	$J_{n-1}$	$P_n$	$P_{n-1}$	$J_{n-1}$	
1	8.609	0.507 <sup>a</sup>	-0.124	0.179	0.649	-0.152	0.179	0.428
2	13.835 <sup>d</sup>	0.462 <sup>a</sup>	0.025	0.321 <sup>c</sup>	0.570	0.031	0.311	0.430
3	15.665 <sup>c</sup>	0.344 <sup>a</sup>	-0.278 <sup>b</sup>	0.465 <sup>a</sup>	0.445	-0.356	0.462	0.378
4	12.132 <sup>d</sup>	0.535 <sup>a</sup>	0.017	0.136	0.699	0.021	0.132	0.525
5	15.936 <sup>b</sup>	0.369 <sup>a</sup>	0.010	0.092	0.624	0.017	0.089	0.402
6	8.956	0.326 <sup>a</sup>	-0.034	0.239 <sup>d</sup>	0.624	-0.065	0.236	0.403
7	5.645	0.362 <sup>a</sup>	-0.089	0.301 <sup>c</sup>	0.663	-0.162	0.296	0.497
8	11.169 <sup>d</sup>	0.241 <sup>a</sup>	-0.099	0.335 <sup>c</sup>	0.438	-0.180	0.336	0.278
9	0.272	0.537 <sup>a</sup>	0.021	0.027	0.766	0.031	0.027	0.595
10	23.845 <sup>a</sup>	0.289 <sup>a</sup>	-0.087 <sup>d</sup>	0.082	0.711	-0.214	0.083	0.549
11	5.666	0.505 <sup>a</sup>	0.032	0.118	0.629	0.040	0.121	0.391
12	23.386 <sup>d</sup>	0.341 <sup>c</sup>	-0.131	0.034	0.304	-0.121	0.034	0.121
13	73.594 <sup>a</sup>	0.354	-0.339	0.139	0.190	-0.179	0.140	0.071
14	20.624 <sup>b</sup>	0.481 <sup>a</sup>	0.012	-0.025	0.670	0.017	-0.025	0.447
15	12.699 <sup>d</sup>	0.862 <sup>a</sup>	-0.120	0.060	0.791	-0.110	0.058	0.635
16	7.017	0.481 <sup>a</sup>	0.008	0.241 <sup>d</sup>	0.721	0.012	0.236	0.624
17	5.544	0.687 <sup>a</sup>	-0.127	0.197 <sup>d</sup>	0.767	-0.150	0.197	0.644
18	9.407 <sup>b</sup>	0.196 <sup>a</sup>	-0.014	0.113	0.649	-0.048	0.115	0.443
19	-0.134	0.466 <sup>a</sup>	0.122	0.071	0.593	0.154	0.073	0.384
20	10.051 <sup>c</sup>	0.495 <sup>a</sup>	0.001	0.105	0.790	0.002	0.104	0.668
21	9.173	0.323 <sup>a</sup>	0.009	0.250 <sup>d</sup>	0.472	0.013	0.261	0.256
22	0.607	0.384 <sup>a</sup>	0.130 <sup>d</sup>	-0.145	0.762	0.246	-0.146	0.575
23	19.059 <sup>a</sup>	0.396 <sup>a</sup>	-0.036	0.036	0.832	-0.073	0.035	0.696
24	5.465	0.318 <sup>a</sup>	-0.026	0.199 <sup>d</sup>	0.511	-0.041	0.200	0.329
25	14.684 <sup>a</sup>	0.334 <sup>a</sup>	-0.046	0.219 <sup>d</sup>	0.647	-0.089	0.238	0.502
26	7.768	0.444 <sup>a</sup>	-0.050	0.227 <sup>d</sup>	0.595	-0.067	0.235	0.398
27	10.631 <sup>d</sup>	0.390 <sup>a</sup>	-0.140 <sup>d</sup>	0.385 <sup>a</sup>	0.633	-0.227	0.383	0.462
28	8.563	0.667 <sup>a</sup>	-0.022	0.123	0.777	-0.025	0.126	0.619
Mean	12.852	0.432	-0.049	0.162	0.626	-0.060	0.162	0.455
SD	13.449	0.142	0.101	0.131	0.150	0.122	0.131	0.155

<sup>a</sup>  $p < .0001$ , <sup>b</sup>  $p < .001$ , <sup>c</sup>  $p < .01$ , <sup>d</sup>  $p < .05$

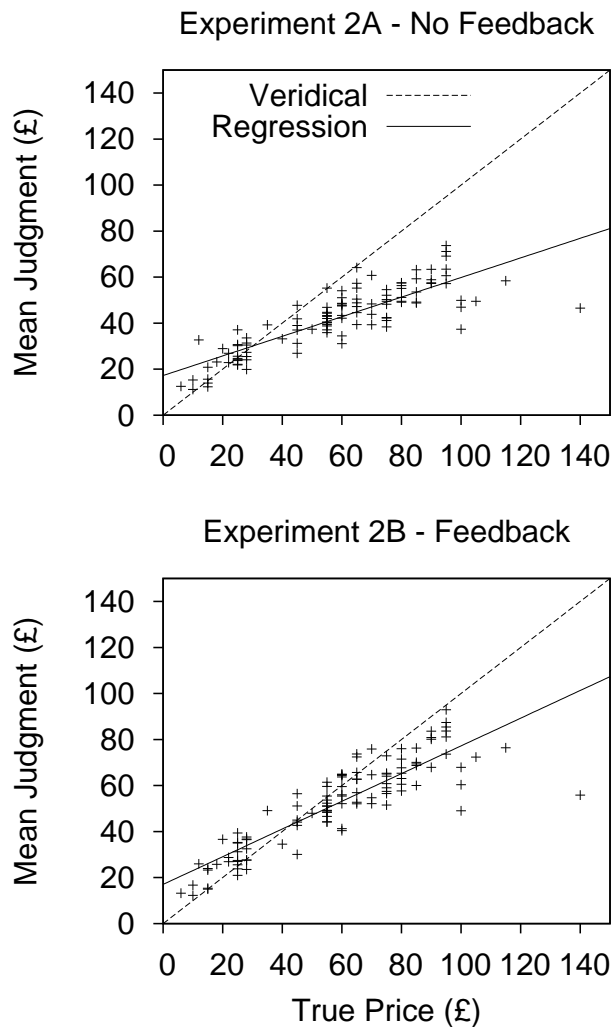


Figure 4: Relationship between true price and judged price in Experiments 2A and 2B.

response assimilation is due to systematic drift over the course of the experiment. We repeated the sequential effects analysis with trial number included as a predictor. The mean coefficient for the trial number term was not significant ( $M = -0.019, SD = 0.075, t(27) = 1.30, p = .204$ ). There was significant assimilation to  $J_{n-1}$ , ( $M = 0.124, SD = 0.138, t(27) = 4.76, p < .001$ , and contrast to  $P_{n-1}$  ( $M = -0.037, SD = 0.103$ ), although the latter effect missed significance ( $t(27) = 1.91, p = .066$ ). The response assimilation we observed therefore does not seem due to systematic drift over the session, although such effects are an important direction for study (Petzold & Haubensak, 2001).

We used independent-samples  $t$ -tests to examine whether participant gender influenced the regression coefficients. The results were not significant for any of the coefficients (for the intercept,  $t(26) = 1.73, p = .096$ ;

for  $P_n, t(26) = 1.05, p = .301$ ; for  $P_{n-1}, t(26) = 1.44, p = .163$ ; for  $J_{n-1}, t(26) < 1$ ).

We examined second order effects using the approach described for Experiment 1. The only modification was that, in keeping with the rest of our analyses for this experiment, we assumed a linear relationship between stimulus price and judgment,  $J_n = \kappa P_n + k$ . We therefore examined the correlation between  $J_n - (\kappa P_n + k)$  and  $J_{n-1} - (\kappa P_{n-1} + k)$ . Similarly, we used  $P_n - P_{n-1}$  (rather than  $\log P_n - \log P_{n-1}$ ) as a measure of jump size. The results are shown in the middle of Figure 3. The inverted v-shape seen in psychophysical studies is apparent and the jump size effect is significant,  $F(4, 108) = 3.54, p = .009, \eta_p^2 = .116$ .

Finally, we tried a regression model which included events from two trials back as predictors. There was no evidence that either predictor influenced the current judgment: For  $P_{n-2}$ , two of the 28 participants had significantly positive coefficients and one had a significant negative coefficient; for  $J_{n-2}$ , one participant had a significant positive coefficient. One sample  $t$ -tests on the mean coefficients similarly revealed no effect of  $P_{n-2}$  ( $M = 0.010, SD = 0.120, t(27) < 1$ ) or  $J_{n-2}$  ( $M = 0.015, SD = 0.110, t(27) < 1$ ).

## 4 Experiment 2B

Experiment 2B was virtually identical to Experiment 2A, except that participants were told the true price of each item after they entered their judgments. Feedback exerts a marked effect on perceptual judgments (e.g., Mori & Ward, 1995), and the effects of feedback on the form of sequential dependencies illuminates the ways in which participants use local context to make their judgments (Stewart et al., 2005).

### 4.1 Method

#### 4.1.1 Participants

Twenty eight new participants were recruited from the same population as Experiment 2A. None had participated in Experiment 2A and all were naive to the purposes of the experiment. Fourteen were males aged 19-25 years ( $M = 20.8, SD = 1.7$ ) and fourteen were females aged 18-33 ( $M = 21.1, SD = 4.2$ ).

#### 4.1.2 Stimuli, Design, and Procedure

The stimuli, design and procedure were identical to Experiment 2A, except that each item's true price was shown for the first 750-ms of the interval between the end of the response window and the presentation of the next item.

## 4.2 Results and Discussion

Participants failed to respond on 2.1% of the test trials, and a few additional trials were excluded as before (0.32% of test trials).

### 4.2.1 Relationship between actual price and judged price

The bottom panel of Figure 4 shows the mean judgment for each product against the item's true price. As in Experiment 2A, we used raw prices and judgments in all analyses. For every participant there was a significant positive relationship between judged price and true price. As in previous experiments, the line relating judgment to true price is swivelled towards the horizontal.

To examine whether the provision of feedback improved the accuracy of judgments, we calculated root mean squared error (RMSE) for each participant in Experiments 2A and 2B. For Experiment 2A, the mean RMSE was 30.02 ( $SD = 9.11$ ); for Experiment 2B, the mean RMSE was 21.53 ( $SD = 3.28$ ). We conducted a 2x2 between subjects ANOVA with condition (Feedback vs. No feedback) and gender (Male vs. Female) as factors. The results indicated a significant effect of condition: the provision of feedback improved accuracy ( $F(1, 52) = 22.2, p < .001, \eta_p^2 = .299$ ). However, there was no main effect of gender ( $F(1, 52) = 3.0, p = .088, \eta_p^2 = .055$ ) and no interaction ( $F(1, 52) < 1$ ).

### 4.2.2 Sequential Effects

We examined sequential effects in the same way as before. The coefficients for each participant are shown in Table 4. For all 28 participants there is a significant positive relationship between true price and judged price. For the  $P_{n-1}$  term, 17 participants have a positive coefficient (6 of which are significant) and 11 have a negative coefficient (2 significant). For the  $J_{n-1}$  term, 14 show a positive relationship (5 significant) and 14 show a negative relationship (none significant).

One-sample  $t$ -tests confirm the pattern suggested by inspection of the individual coefficients. There is a significant positive effect of true price,  $t(27) = 27.6, p < .001$ , a significant positive effect of  $P_{n-1}$ ,  $t(27) = 2.27, p = .031$ , and no effect of  $J_{n-1}$ ,  $t(27) = 1.84, p = .077$ . Inclusion of trial number as a predictor made no difference to this pattern.

The second order effects are shown in Figure 3. There is some indication of an inverted v-shape, but the effect of jump size is not significant ( $F(4, 108) = 1.10, p = .360, \eta_p^2 = .039$ ).

Inclusion of  $P_{n-2}$  and  $J_{n-2}$  as predictors in the regression equation indicated no consistent effect of the

events from two trials back in the sequence. (One participant showed a significant positive  $P_{n-2}$  coefficient; one showed a significant negative  $J_{n-2}$  coefficient; one sample  $t$ -tests on the mean coefficients showed no effect of either  $P_{n-2}$ ,  $M = 0.014, SD = 0.086, t(27) < 1$  or  $J_{n-2}$ ,  $M = -0.001, SD = 0.144, t(27) < 1$ .)

Independent samples  $t$ -tests indicated no effect of gender on any of the coefficients (for the intercept,  $t(26) = 1.06, p = .300$ ; for the  $P_n$  predictor,  $t(26) = 1.40, p = .172$ ; for  $P_{n-1}$ ,  $t(26) < 1$ ; for  $J_{n-1}$ ,  $t(26) = 1.08, p = .289$ ).

Comparison of the  $R^2$  changes listed in Table 2 shows that feedback increased the dependence on  $P_{n-1}$  and reduced the dependence on  $J_{n-1}$ . However, feedback did not appreciably diminish the total effect of the events from the previous trial.

In short, the provision of feedback shifts the effect of the preceding stimulus from contrast to assimilation and reduces the effect of the preceding judgment.

### 4.2.3 Effect of feedback on sequential effects

In order to compare the feedback and no-feedback conditions directly, and to see whether the feedback manipulation interacted with gender, we conducted a series of 2x2 between-subjects ANOVAs with condition (Feedback vs. No feedback) and gender (Male vs. Female) as factors. We conducted a separate ANOVA for each of the predictors in the regression model, using the regression coefficients as the dependent variables.

For the intercept, there was no effect of condition,  $F(1, 52) < 1$ , but a significant effect of gender,  $F(1, 52) = 4.1, p = .049, \eta_p^2 = .072$ , with the intercept for male participants ( $M = 14.83, SD = 13.09$ ) larger than that for females ( $M = 9.40, SD = 5.58$ ). There was no interaction between condition and gender,  $F(1, 52) = 1.27, p = .265, \eta_p^2 = .024$ .

For the true price  $P_n$ , there was a significant effect of condition,  $F(1, 52) = 24.3, p < .001, \eta_p^2 = .319$ ; as can be seen by comparison of Tables 3 and 4, the mean coefficient became larger in the Feedback condition. As described above, the relationship between judged price and true price was too shallow in both conditions. The finding that feedback rendered the slope steeper is therefore consistent with the results of the RMSE analysis which showed that feedback improved the accuracy of judgments. The relationship between true price and judgment was not influenced by gender,  $F(1, 52) = 2.89, p = .095, \eta_p^2 = .053$ , and there was no interaction between condition and gender,  $F(1, 52) < 1$ .

The effect of the preceding item's true price,  $P_{n-1}$ , was significantly influenced by the provision of feedback,  $F(1, 52) = 11.2, p = .002, \eta_p^2 = .177$ . As indicated above, feedback shifted the effects of the preced-

Table 4: Regression coefficients for Experiment 2B (Feedback provided).

Participant	Unstandardized				Standardized			$R^2$
	int	$P_n$	$P_{n-1}$	$J_{n-1}$	$P_n$	$P_{n-1}$	$J_{n-1}$	
1	10.379	0.689 <sup>a</sup>	0.159	-0.078	0.670	0.155	-0.079	0.470
2	16.262 <sup>d</sup>	0.421 <sup>a</sup>	0.219 <sup>d</sup>	0.053	0.462	0.239	0.052	0.325
3	11.218 <sup>d</sup>	0.480 <sup>a</sup>	0.103	-0.036	0.730	0.153	-0.038	0.541
4	15.292 <sup>d</sup>	0.668 <sup>a</sup>	-0.043	-0.025	0.696	-0.043	-0.025	0.484
5	12.187 <sup>d</sup>	0.662 <sup>a</sup>	0.104	-0.086	0.793	0.125	-0.088	0.616
6	28.676 <sup>a</sup>	0.393 <sup>a</sup>	0.135	-0.166	0.474	0.163	-0.166	0.260
7	5.451	0.571 <sup>a</sup>	0.140	0.137	0.754	0.186	0.137	0.616
8	8.439	0.719 <sup>a</sup>	0.121	0.079	0.738	0.126	0.079	0.606
9	13.862 <sup>d</sup>	0.611 <sup>a</sup>	-0.226 <sup>d</sup>	0.311 <sup>c</sup>	0.699	-0.256	0.306	0.567
10	12.766	0.304 <sup>a</sup>	0.216 <sup>c</sup>	0.149	0.437	0.323	0.147	0.305
11	5.278	0.704 <sup>a</sup>	-0.083	0.232 <sup>c</sup>	0.690	-0.080	0.261	0.510
12	16.359 <sup>c</sup>	0.520 <sup>a</sup>	0.105	-0.010	0.701	0.142	-0.010	0.506
13	11.515	0.696 <sup>a</sup>	-0.126	0.215 <sup>d</sup>	0.695	-0.125	0.220	0.519
14	8.577	0.554 <sup>a</sup>	-0.086	0.242 <sup>d</sup>	0.743	-0.118	0.245	0.545
15	8.067	0.458 <sup>a</sup>	0.022	0.232 <sup>d</sup>	0.536	0.025	0.225	0.337
16	7.937	0.719 <sup>a</sup>	0.084	-0.041	0.782	0.094	-0.041	0.621
17	8.762	0.768 <sup>a</sup>	0.200 <sup>d</sup>	-0.146	0.801	0.226	-0.150	0.666
18	11.983 <sup>d</sup>	0.591 <sup>a</sup>	0.217 <sup>d</sup>	-0.035	0.702	0.250	-0.035	0.549
19	8.452	0.665 <sup>a</sup>	0.060	-0.002	0.782	0.071	-0.002	0.614
20	6.642	0.613 <sup>a</sup>	-0.008	0.092	0.794	-0.011	0.091	0.629
21	12.784 <sup>d</sup>	0.655 <sup>a</sup>	-0.033	-0.063	0.758	-0.039	-0.064	0.586
22	19.085 <sup>c</sup>	0.568 <sup>a</sup>	-0.053	0.052	0.640	-0.060	0.052	0.405
23	2.070	0.502 <sup>a</sup>	0.214 <sup>c</sup>	0.112	0.621	0.270	0.111	0.560
24	10.955 <sup>d</sup>	0.570 <sup>a</sup>	-0.041	0.196	0.763	-0.054	0.196	0.614
25	18.022 <sup>b</sup>	0.710 <sup>a</sup>	-0.023	-0.012	0.810	-0.026	-0.013	0.649
26	21.302 <sup>a</sup>	0.698 <sup>a</sup>	-0.191 <sup>d</sup>	0.091	0.769	-0.208	0.091	0.583
27	7.174	0.731 <sup>a</sup>	0.136	-0.057	0.692	0.131	-0.059	0.514
28	-0.483	0.585 <sup>a</sup>	0.299 <sup>a</sup>	-0.154	0.750	0.381	-0.154	0.600
lex $M$	11.393	0.601 <sup>a</sup>	0.058	0.046	0.696	0.073	0.046	0.528
$SD$	6.000	0.115 <sup>a</sup>	0.135	0.132	0.103	0.160	0.134	0.110

<sup>a</sup>  $p < .0001$ , <sup>b</sup>  $p < .001$ , <sup>c</sup>  $p < .01$ , <sup>d</sup>  $p < .05$

ing price from a negative dependency (contrast) to a positive one (assimilation). There was no effect of gender,  $F(1, 52) = 1.01$ ,  $p = .320$ ,  $\eta_p^2 = .019$ , and no interaction,  $F(1, 52) < 1$ .

Finally, the effect of the preceding judgment was significantly influenced by the provision of feedback,  $F(1, 52) = 10.7$ ,  $p = .002$ ,  $\eta_p^2 = .171$ ; the large assimilation to  $J_{n-1}$  found in Experiments 1 and 2A is essentially eliminated by the provision of feedback. There was no effect of gender,  $F(1, 52) = 1.26$ ,  $p = .266$ ,  $\eta_p^2 = .024$ , and no interaction,  $F(1, 52) < 1$ .

## 5 General discussion

In three experiments we have found sequential dependencies in judgments of price which match those seen in psychophysical tasks. In Experiment 1 and Experiment 2A, the judged price of the current item assimilated towards the judgment made about the previous item but contrasted away from the true price of that item. This parallels the finding in studies of magnitude estimation, cross-modality matching, category judgment and absolute identification (e.g., Jesteadt et al., 1977; Matthews & Stewart, in press; Mori, 1998; Mori & Ward, 1995; Petzold & Haubensak, 2001; Ward, 1987). In Experiment 2B, we found that telling people the true price of each item after they entered their judgment improved accuracy and changed the pattern of sequential dependencies; contrast to the preceding item's true price was replaced by assimilation, and the assimilation to the preceding judgment largely disappeared. These findings have a number of implications for our understanding of how people make a series of judgments, which we now discuss in turn.

### 5.1 The effect of the previous judgment

In the absence of feedback, the current judgment assimilates towards the previous one. There are several possible interpretations of this result; some have suggested that when the participant is unsure of her judgment she simply repeats the previous response (e.g., Garner, 1953; see also Treisman & Williams, 1984). However, studies of perceptual identification suggest that there is genuine assimilation rather than mere repetition (e.g., Stewart et al., 2005), and a number of psychophysicists have interpreted response assimilation as indicating that the previous item is used as a point of reference for the current judgment. That is, rather than evaluating each item with respect to a long-term set of referents or a fixed internal scale, each judgment proceeds at least partly by comparison with the previous item.

Exactly how this works is a topic of debate (e.g., DeCarlo & Cross, 1990; Laming, 1984; Stewart et al., 2005), but Laming (1995) provides a concise statement of the general principal in a discussion of cervical smear tests: "Suppose now that one smear, call it  $S_n$ , is diagnosed as positive. The next smear,  $S_{n+1}$ , is compared with  $S_n$  and, if it is judged to be more abnormal than  $S_n$  or about the same, is also diagnosed as positive. This means that the mere fact of  $S_n$  being called 'positive' increases the likelihood that  $S_{n+1}$  will be classified 'positive' too" (Laming, 1995, p. 513). Laming's article is a (non-experimental) attempt to make sense of the failure of a professional cytologist to identify a large number of cancerous cervical smears. The response assimilation seen in our data (when feedback is absent) suggests that a similar use of local context applies when people make judgments of price, and bolsters the idea that this is a general principal of human judgment not restricted to situations involving highly confusable, unidimensional stimuli.

### 5.2 The effect of the previous stimulus

In Experiments 1 and 2A, where feedback was absent, we found weak contrast to the true price of the previous item. A similar finding has been reported in studies of magnitude estimation (e.g., Jesteadt et al., 1977) and perceptual identification (e.g., Mori & Ward, 1995). The role of the previous stimulus in shaping psychophysical judgments has been given a number of interpretations. Notably, several of these explanations seem to be restricted to judgments of single physical aspects of unidimensional stimuli (e.g., Brown et al., 2008). For example, Ward (1979) argues that the central nervous system forms an internal representation of the stimulus and that "because of the excitatory-center inhibitory-surround nature of such neural representations . . . the center of the internal representation of the stimulus on Trial  $n$  is moved away from (contrasted with) its centre on Trial  $n - 1$ " (Ward, 1979, p. 446).

Such accounts are difficult to adapt to judgment situations where the to-be-judged items are complex multidimensional objects and the judgment dimension is not a simple physical property. Alternative explanations for the effects of the preceding stimulus, which assume confusion of successive items in memory, seem more applicable to the current scenario (e.g., Lockhead & King, 1983). Such memory-based accounts have implications for the effects of manipulating the time between successive items and can be tested by manipulating the inter-judgment interval (DeCarlo, 1992; Matthews & Stewart, in press).

Although on average participants in Experiments 1 and 2A showed contrast to  $P_{n-1}$ , some individuals showed weak assimilation. We believe this is most likely due to

noise. However, Beckstead (2008) has noted that when participants make judgments of multi-dimensional stimuli, some dimensions produce contrast and some produce assimilation. The heterogeneity of  $P_{n-1}$  coefficients might imply that participants are primarily basing their decisions on a dimension which elicits contrast whilst others focus on a dimension which elicits assimilation.

### 5.3 The effects of providing feedback

In Experiment 2B, telling participants the true price of each item after they entered their judgment substantially altered their performance. As would be expected from psychophysical studies, the provision of feedback improved accuracy (e.g., Ward & Lockhead, 1971). In addition, feedback largely eliminated the assimilation to  $J_{n-1}$  and shifted the effects of  $P_{n-1}$  from weak contrast to weak assimilation. These results are consistent with the idea that participants use the previous judgment as a point of reference. In the absence of feedback, the judgment process is presumably something like: "This item looks like it costs a bit more than the last one. I said the last one would cost £100, so I'll say this one costs £120." When feedback is provided, the decision is likely to be made with reference to that information: "This one looks a bit more expensive than the last, and I was told that that one cost £100 . . ."

Although Experiment 2B indicated assimilation to the true price of the preceding item, the mean coefficient was relatively small when compared to that for  $J_{n-1}$  in Experiments 1 and 2A. This is presumably because the perceptual or mnemonic effects responsible for contrast to  $P_n$  were still present, such that the net influence of  $P_{n-1}$  was a combination of weak contrast (of the type seen in Experiments 1 and 2A) driven by some perceptual or memory factor, and strong assimilation, driven by comparative judgment. It is also noticeable in Table 2 that the provision of feedback did not reduce the overall effect of the local context (cf. Mori & Ward, 1995).

Although feedback reduced the net effect of  $J_{n-1}$ , several participants still show significant response assimilation, perhaps because of a tendency to repeat the most recent judgment. The heterogeneity of regression coefficients in Experiment 2B may indicate that the effects of feedback differ from participant to participant. More generally, individuals may differ in the judgment strategies they adopt and the sequential effects that result.

### 5.4 The effects of expertise and amount of information

Our attempt to investigate the effects of expertise in Experiments 2A and 2B yielded little. This was almost cer-

tainly due to a failure of the manipulation; we had conjectured that female participants would know more about women's shoes than male participants would. In post-experimental debriefing we asked the participants if they had ever been to a Topshop store or visited the Topshop website. All 28 female participants reported having visited the store or website; however, 18 of the male participants reported having been, too (10 in Experiment 2A, 8 in Experiment 2B; note that Topshop exclusively sells women's clothing). Thus, although the female participants in our experiments were probably more knowledgeable about the items being judged than were the males, the difference was relatively small.

In general we would expect expertise and information about the items presented for judgment to influence sequential dependencies. In magnitude estimation experiments, decreasing stimulus information (for example by presenting the stimuli for less time) increases the dependence on the previous response (Ward, 1979). However, this effect depends upon whether or not feedback is provided. In absolute identification tasks with feedback provided, decreasing stimulus information decreases dependency on the preceding response but increases the dependency on the previous stimulus (see e.g., Lockhead, 1984). One can envisage that, when the task is difficult, the participant uses the events from the previous trial when making her judgment. If she is provided with feedback, she elects to use this information, rather than her own uncertain judgment, as the point of reference. These effects of feedback and uncertainty will be important in reducing sequential biases in real-world applications.

### 5.5 Choice of regression equation and depth of the effects

We used Equation 1 to assess local context effects because it has been widely used in psychophysical research and because it provides readily-interpretable results. There are, of course, other ways to test for sequential dependencies. For example, Beckstead (2008) has recently applied time-series analysis to expert medical judgments with multiple predictor cues. Such analyses hold great promise; however, the simple regression analysis employed here has the advantage that it allows direct comparison with previous work in psychophysical studies, providing a theoretical framework for interpreting the results and a testable set of predictions regarding the effects of experimental manipulations such as the provision of feedback or changes in stimulus information.

DeCarlo and Cross (1990; DeCarlo, 1992) have suggested replacing the  $J_{n-1}$  predictor in Equation 1 with an autocorrelated error term, such that it is the error in judgment, rather than raw responses, which are correlated over successive trials. In magnitude estimation studies,

the results of this analysis are usually similar to those obtained with Equation 1 with the exception that contrast to the preceding stimulus is replaced by assimilation, which has implications for the precise interpretation of the perceptual/mnemonic component of the sequential dependencies. However, the finding that successive responses are correlated because the previous trial provides a point of reference for the current judgment is retained in DeCarlo and Cross's (1990) analysis.

We found no evidence for effects of stimuli or responses from two trials previously, (although we should be cautious about accepting this null result). Psychophysical studies have similarly found that in magnitude estimation tasks — where, as in the current experiments, the participant is free to give any number she likes as a response — only the immediately preceding trial influences judgment (e.g., Jesteadt et al., 1977; Petzold & Haubensak, 2001), whereas in categorization and identification tasks — where the response set is constrained — the effects extend for two or more trials (Petzold & Haubensak, 2001; Staddon, King, & Lockhead, 1980). Petzold and Haubensak (2001) have suggested that this difference occurs because magnitude estimation involves comparing the stimulus with a single referent (the previous item), whereas category judgment involves gauging the position of the stimulus in a subjective range defined by two endpoints. These ideas could be tested by replacing the price judgment task used in our experiments with a price categorization task and seeing whether deeper sequential effects emerge.

## 5.6 Second order effects

We found some evidence that the correlation between successive responses was greater when successive stimuli were closer together, particularly in Experiment 2A (see Figure 3), although the small number of trials and binning of different jump sizes mean that this result must be treated with caution. In psychophysics, this result has been given a number of interpretations (e.g., Ward, 1979). In particular, Laming (1984) has taken it as evidence that participants judge each stimulus with respect to the last, but that their judgment of stimulus differences is very imprecise, to the point that judgments are little better than ordinal. That is, participants rate each stimulus as “a lot less,” “a little less,” “about the same,” “a bit more” or “a lot more” than the previous item.

## 5.7 The central tendency of judgment

In all three experiments, we found that the regression lines relating participants' judgments to the true prices were swivelled towards the horizontal: Judgments were pulled towards the centre of the range, with an overesti-

mation of cheap products and an underestimation of expensive ones. If this reflects a genuine central tendency of judgment (rather than just regression to the mean) then the standard deviation of the responses for each participant will be less than the standard deviation of the true prices. (We are grateful to Jonathan Baron for this suggestion.) For Experiment 1, the *SD* of the (log) judgments was smaller than *SD* of the (log) prices for all 25 participants (for judgments,  $M = 0.714$ ,  $SD = 0.141$ ; for true prices,  $M = 0.975$ ,  $SD = 0.002$ ; note that the small number of excluded trials explains why there is some slight between-participant variation in the standard deviation of the true prices.) A paired-samples *t*-test indicated that the difference is significant,  $t(24) = 9.26$ ,  $p < .001$ . Similarly, in Experiment 2A the *SD* for (untransformed) judgments was smaller than that for (untransformed) prices for 25 of the 28 participants (for judgments,  $M = 19.85$ ,  $SD = 7.82$ ; for true prices,  $M = 27.24$ ,  $SD = 0.44$ ,  $t(27) = 4.99$ ,  $p < .001$ ). In Experiment 2B, the same pattern was found (24 of the 28 participants had smaller judgment *SD* ( $M = 23.66$ ,  $SD = 3.21$ ) than true price *SD* ( $M = 27.19$ ,  $SD = 0.54$ ;  $t(27) = 5.80$ ,  $p < .001$ ). As one would expect from the foregoing discussion of feedback effects, the difference between the judgment *SD* and price *SD* was greater in the absence of feedback ( $M = 7.39$ ,  $SD = 7.84$ ) than when feedback was provided ( $M = 3.53$ ,  $SD = 3.22$ ;  $t(27) = 2.41$ ,  $p = .021$ ).

This central tendency of judgment has long been recognized by psychophysicists (e.g., Hollingworth, 1909); in studies of magnitude estimation, it is referred to as the “regression effect” (e.g., Reynolds & Stevens, 1960; Stevens & Guirao, 1962). Laming (2004) has argued that this kind of bias is common in real-world judgments, and Garner (1953) has raised the possibility that this central tendency is itself a consequence of sequential effects. The central tendency can also be thought of as a rational strategy: If a person is completely uncertain, it makes sense to guess the average value; in a state of partial uncertainty, it may be sensible to bias one's judgment towards the mean.

Establishing the validity of these ideas in the current context by systematically manipulating the range of items presented for judgment would be a useful direction for future work.

## 5.8 Relationship to anchoring

We have argued that participants use the local context as a frame of reference and have related our findings to those from psychophysical research. However, the current results also connect to a quite different body of research, namely that concerned with anchoring. Anchoring occurs when people's judgments are biased towards some extraneous value. In some situations the anchor is self-

generated, in which case people proceed by a process of “anchor and adjustment” (Tversky & Kahneman, 1974). For example, when asked the year in which George Washington was elected president, most people begin by thinking of the year of independence and then adjusting away from that value (Epley & Gilovich, 2001). In other situations the anchor is provided by the experimenter. For example, Tversky & Kahneman (1974) asked participants to judge whether the percentage of African countries in the United Nations (UN) is higher or lower than some value (the anchor). Subsequent estimates of the percentage of African countries in the UN were biased towards the anchor. Such experimenter-provided anchors seem to influence judgment by activating anchor-consistent knowledge (Mussweiler & Strack, 1999), and influence judgments even when the anchor is not relevant to the task (Wilson, Houston, Etling, & Brekke, 1996) or is subliminally presented (Mussweiler & Englich, 2005).

There is a formal similarity between anchoring and the sequential effects found in both the current experiments and in psychophysical investigations. On each trial of a magnitude estimation experiment, the participant has recently encountered a number (their own previous response) which may serve as an anchor for the current response. When feedback is provided the most recently encountered number will be the feedback from the previous trial, which may again serve as an anchor and mask the anchoring on the previous judgment. Moreover, the anchoring effect depends upon the participant’s knowledge of the topic about which they are asked (Wilson et al., 1996) in much the same way that increasing information about a stimulus reduces the magnitude of sequential effects (Ward, 1979).

The idea that anchoring and sequential dependencies share common mechanisms leads to novel empirical directions. For example, self-generated anchors produce greater effects when the participant is engaged in acceptance behaviour (head nodding) than rejection behaviour (head shaking), presumably because head shaking decreases the chance that a given adjustment from the anchor will be accepted as the correct answer (Epley & Gilovich, 2001). If sequential dependencies arise because the previous response serves as a self-generated anchor then the head nodding/shaking manipulation ought to influence the magnitude of response assimilation. Clearly, this effect is not overtly predicted by psychophysical accounts of sequential dependencies.

## 5.9 Conclusions

The immediate context affects judgments of price in much the same way as it affects judgments of loudness or brightness, and it seems likely that in both cases participants partially use the most recent item as a point of

reference for the current judgment. The magnitude and form of this dependency can be reduced by the provision of feedback and, most likely, by expertise (although the latter effect was not found here). These results have implications for our understanding of judgment in both real-world and psychophysical settings. They are also of practical importance because biases introduced by local context necessarily impair judgment accuracy. Finally, they suggest connections between hitherto entirely separate domains: sequential effects in psychophysical experiments and anchoring in decision-making research.

## References

- 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.
- Beckstead, J. W. (2007). A note on determining the number of cues used in judgment analysis studies: The issue of type II error. *Judgment and Decision Making*, 2, 317–325.
- Beckstead, J. W. (2008). Modeling sequential context effects in judgment analysis: A time series approach. *Judgment and Decision Making*, 3, 570–584.
- Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and responses times in absolute identification. *Psychological Review*, 115, 396–425.
- 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., & Cross, D. V. (1990). Sequential effects in magnitude scaling: models and theory. *Journal of Experimental Psychology: General*, 119, 375–396.
- Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17, 311–318.
- Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, 46, 373–380.
- Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, 3, 409–414.
- Hollingworth, H. L. (1909). The inaccuracy of movement. *Archives of Psychology*, 13.
- Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 92–104.



- Laming, D. (1984). The relativity of 'absolute' judgments. *British Journal of Mathematical and Statistical Psychology*, 37, 152–183.
- Laming, D. (1995). Screening cervical smears. *British Journal of Psychology*, 86, 507–516.
- Laming, D. (2004). *Human Judgment: The eye of the beholder*. London: Thomson Learning.
- 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.
- Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 461–473.
- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 149–157.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203–220.
- Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review*, 14, 989–993.
- Matthews, W. J., & Stewart, N. (in press). The effect of inter-stimulus interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology*.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for information processing. *Psychological Review*, 63, 81–97.
- Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology*, 52, 72–83.
- Mori, S., & Ward, L. M. (1995). Pure feedback effects in absolute identification. *Perception & Psychophysics*, 57, 1065–1079.
- Mussweiler, T., & Englich, B. (2005). Subliminal anchoring : Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes*, 98, 133–143.
- Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm : A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136–164.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd Ed.). London : Irwin.
- Parducci, A. (1965). Category judgment : A range frequency model. *Psychological Review*, 72, 407–418.
- Parducci, A. (1968). The relativism of absolute judgment. *Scientific American*, 219, 84–90.
- Petzold, P., & Haubensak, G. (2001). Higher order sequential effects in psychophysical judgments. *Perception & Psychophysics*, 63, 969–978.
- Reynolds, G. S., & Stevens, S. S. (1960). Binaural summation of loudness. *The Journal of the Acoustical Society of America*, 32, 1337–1344.
- Staddon, J. E. R., King, M., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 290–301.
- Stevens, S. S., & Guirao, M. (1962). Loudness, reciprocity, and partition scales. *The Journal of the Acoustical Society of America*, 34, 1466–1471.
- Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, 112, 881–911.
- Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. In N. Chater & M. Oaksford (Ed.s), *The probabilistic mind: Prospects for Bayesian cognitive science*, (pp. 261–276). Oxford: Oxford University Press.
- Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91, 68–111.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Vlaev, I. & Chater, N. (2007). Context effects in games: Local versus global sequential effects on choice in the prisoner's dilemma game. *Judgment and Decision Making*, 2, 380–389.
- Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 444–459.
- 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 intramodality 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. (1990). Mixed-methods mixed-modality psychophysical scaling. *Perception & Psychophysics*, 48, 571–582.
- Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, 84, 27–34.

- Ward, L. M., & Lockhead, G. R. (1971). Response system processed in absolute judgment. *Perception & Psychophysics*, *9*, 73–78.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387–402.