Relative judgement is relatively difficult: Evidence against the role of relative judgement in absolute identification

Duncan Guest, James. S. Adelman, & Christopher Kent

Nottingham Trent University, University of Warwick, University of Bristol

5940 words

Corresponding Author

Duncan Guest

duncan.guest@ntu.ac.uk

Nottingham Trent University

Burton Street, Nottingham, UK

NG1 4BU

Tel: +44(0)115 84 82701

Abstract

A variety of processes have been put forward to explain absolute identification performance. One difference between current models of absolute identification is the extent to which the task involves accessing stored representations in long term memory (e.g., exemplars in memory, Kent & Lamberts, 2005) or relative judgement (comparison of the current stimulus to the stimulus on the previous trial, Stewart, Brown & Chater, 2005). In two experiments we explored this by tapping into these processes. In Experiment 1 participants completed an absolute identification task using eight line lengths whereby a single stimulus was presented on each trial for identification. They also completed a matching task aimed at mirroring exemplar comparison in which eight line lengths were presented in a circular array and the task was to report which of these matched a target presented centrally. Experiment 2 was a relative judgement task and was similar to Experiment 1 except that the task was to report the difference (jump-size) between the current stimulus and that on the previous trial. The absolute identification and matching data showed clear similarities (faster and more accurate responding for stimuli near the edges of the range and similar stimulus-response confusions). In contrast relative judgment performance was poor suggesting relative judgement is not straightforward. Moreover, performance as a function of jump-size differed considerably between the relative judgement and absolute identification tasks. Similarly, in the relative judgement task, predicting correct stimulus identification based on successful relative judgement yielded the reverse pattern of performance observed in the absolute identification task. Overall the data suggest that relative judgement does not underlie absolute identification and that the task is more likely reliant on an exemplar comparison process.

Absolute identification is a simple task in which multiple stimuli (N > 2) vary along a single psychophysical dimension (e.g., line length). Each stimulus is paired with a numerical label referring to its rank magnitude in the set. On each trial a stimulus is presented and the task is to respond with its label. Correct feedback is then typically given. Absolute identification was first examined by Cohn (1900; for an early history see Wever & Zener, 1928) and has remained of interest due to the surprisingly severe performance limitations participants experience when N > 5 (e.g., Cowan, 2001; Miller, 1956) and because the simple nature of the task means that the processes involved in it are likely to underlie many other cognitive tasks (Brown et al, 2008, 2009; Dodds et al, 2011a, 2011b; Donkin et al, 2009; Guest, Kent & Adelman, 2010, 2015; Matthews & Stewart, 2009; Stewart & Matthews, 2009).

There are a wide range of robust effects in absolute identification in terms of choice and response time (RT). Extensive reviews of these are given elsewhere (Brown et al 2008; Donkin et al, 2015; Lacouture, 1997; Stewart, Brown & Chater, 2005) but the standard effects that have informed modelling include the following. Performance improvements (in accuracy and RT) are observed for stimuli nearer the edges of the range (the bow effect), as N reduces (the set size effect; Kent & Lamberts, 2005; Lacouture & Marley, 2004) and as the spacing between stimuli increases (the spacing effect; e.g., Stewart, Brown & Chater, 2005). Reducing the stimulus range increases accuracy (range effect; Lockhead & Hinson, 1986) and as N increases the amount of information that can be used asymptotes at just over two bits of information (information transmission limit; Garner, 1953, Pollack, 1952). Sequential effects (e.g., Ward & Lockhead, 1970) typically manifest as: i) assimilation, responses to the stimulus on the current trial tend to be more similar to the stimulus on the previous trial and ii) contrast, responses to the stimulus on the current trial tend to be less similar to the stimulus

shown two trials back than would be expected due to chance, with the effect lasting for up to five trials back .

A number of detailed mathematical models have been developed to account for these effects that differ markedly in the underlying processes hypothesised (see Donkin et al., 2015, for a review). One of the key differences, and the focus of this paper, is the assumption regarding what is represented in the task. Exemplar models assume that there is a long term representation of each stimulus in the set and that absolute identification involves comparing a representation of the current stimulus to these exemplars in memory (e.g., Kent & Lamberts, 2005, Nosofsky, 1997). Capacity limited models posit restricted capacity processes in attention and memory, such as a limited capacity selective attention stage that maintains the experimental context in short term memory via rehearsal (Marley & Cook, 1984, 1986). This context can comprise the ends of the stimulus range (the anchor points, see e.g., Petrov & Anderson, 2005) as well as particular parts of the stimulus range recently attended. This information is then used to make a decision, for example by mapping the stimulus magnitude onto this experimental context (Lacouture & Marley, 1995, 2004). Relative judgement models assume no long term representation of the stimuli. Rather absolute identification performance is determined by making a relative judgement on each trial based on the stimuli seen in recent trials and responses made to these stimuli (The Relative Judgement Model, RJM, Stewart et al., 2005). Some models are hybrids of these general approaches, such as the Selective Attention Mapping and Ballistic Accumulator Model (SAMBA; Brown et al., 2008). SAMBA can account for a range of effects in absolute identification by assuming a long term reference for the experimental context and limited capacity selective attention. However, in order to capture the extent that performance improves when a stimulus is repeated (Rouder et al., 2004) it is necessary to include a relative mechanism, such that on

some proportion of trials, instead of using the end anchors to estimate the stimulus magnitude, participants replace one of the end anchors with the previous stimulus (Brown et al, 2005).

To summarise, models of absolute identification differ considerably in their assumptions about what is represented in the task. Exemplar models assume long term representations, either of the stimuli or of the psychological dimension, limited capacity models assume a representation of the experimental context is held either in short term or long term memory and relative judgement models assume that only representations of the previous stimuli and responses are available. Hybrid models combine the latter two assumptions. These differences are evident in current models of the task which account for a range of choice proportion and RT data (Brown, et al, 2008; Kent and Lamberts, 2005; Lacouture & Marley, 2004; Petrov and Anderson, 2005; and Stewart, et al., 2005). Out of these current models SAMBA (Brown et al., 2008) accounts for the most comprehensive range of findings (Brown et al., 2008; Guest, Kent, & Adelman, 2015) followed by Stewart et al.'s (2005) RJM. Both SAMBA and RJM provide good accounts of the set size effect, bow effect, information transmission limit, spacing effects, range effects and sequential effects, although the RJM does not account for RTs. In third place lie exemplar models including the extended generalised context model (EGCM-RT, Kent & Lamberts, 2005; Guest, Kent & Adelman, 2015) and the exemplar based random walk model (EBRW, Nosofsky & Palmeri, 1997; Nosofsky, 1997) which account for bow effects, set size effects, and some spacing effects in both accuracy and RT. Critically, exemplar models do not yet capture sequential effects. Although there are key differences between the EGCM-RT and the EBRW they share the critical assumption that stimuli are represented in long term memory.

While comparing models in terms of the range of effects they capture is informative, it provides only indirect support for the underlying representations hypothesised, particularly when not all models have been fit to the same data sets to account for the same effects. It is therefore not clear, even though SAMBA captures the most effects, whether a long term representation of the experimental context exists. More direct evidence comes from experimental tests of model assumptions. Stewart et al. (2005) used a methodology in which false feedback was given on some trials. The RJM explicitly outlines how feedback is utilised and so can account for false feedback effects. It was argued that other models not incorporating feedback would struggle to capture these effects. Subsequently however, Brown et al. (2008) showed that these effects could be captured by SAMBA by assuming that feedback helps participants maintain the referents (the end anchors) in memory.

Spacing effects have also been used to try to discriminate models. Brown et al. (2009) argued that the RJM needs to contain some absolute knowledge about the spacing between stimuli in the range in order to account for the effects of unequal stimulus spacing. Stewart and Matthews (2009) therefore proposed an augmented RJM to account for unequal spacing effects which contained knowledge of the differences between all stimuli in the stimulus set. As noted by Donkin et al (2015), relative judgement theorists argue that this is relative knowledge, whereas absolute theorists argue that this represents a long term memory for the structure of the stimulus set.

The time course of sequential effects has also been proposed as a method for discriminating models. In many models sequential effects are caused by processes that should decay over time. In the RJM, a judgement is made relative to the previous item, but the estimate of difference is contaminated by previous estimates of differences. Increasing the inter-trial

interval (ITI) should therefore decrease this interference which in the RJM is the mechanism for producing assimilation and contrast. In SAMBA increasing ITI enables more time for accumulators associated with different responses to decay and return to baseline. The exception is that selective attention to the experimental context maintains activation at the location of the previous stimulus. This results in reduced assimilation but increased contrast. Supporting SAMBA but contradicting the RJM, increasing ITI appears to increase contrast (Donkin, Chan &Tran, 2015; Mathews & Stewart, 2009). However, assimilation appears to be unaffected by ITI. Donkin, Chan, and Tran (2015) suggested a modification to SAMBA to account for this, whereby assimilation is produced instead by adjusting response thresholds to make responses more similar to those on the previous trial. If threshold adjustments are proportional to the distance from the previous response, this mechanism is functionally equivalent to SAMBAs decay mechanism, enabling it to produce assimilation. Crucially there is no reason to assume that threshold changes (and thus assimilation) should change with ITI. However Donkin et al. (2015) note that this modification to SAMBA entails that participants adjust their response thresholds faster than is traditionally assumed in decision making models (e.g., Donkin, Brown, & Heathcote, 2011). The threshold adjustment may therefore only be feasible if it is made before decision making is underway on a given trial.

Currently, based on the range of effects captured and experiments attempting to delineate model predictions, there appears to be slightly more evidence against relative judgement and for some form of absolute memory contributing to the performance limit in absolute identification. However, the matter remains contentious. Interestingly the RJM has been cited much more than other models of absolute identification (some 174 times according to Google Scholar on 25/08/15). This is probably because the idea of relative judgement appears intuitive and simple. Indeed, Brown et al. (2009) note that it is more parsimonious because it

assumes that only a limited set of information is represented. However, this line of reasoning is questionable in terms of what we know about memory. At the very least relative judgement requires maintenance of a short term representation of the stimulus magnitude and feedback from the previous trial. This information has to be impervious to interference from other events including output interference, feedback, onset of a fixation cue, and presentation of a new stimulus. The difference between the current stimulus and the previous stimulus then has to be calculated, with information about previous feedback also being integrated. Add to this maintaining information about stimuli two or three trials back, then it is clear that, even assuming no decay in short term memory, there are a large number of potential causes of interference that the participant must overcome. In comparison, exemplar models (Kent & Lamberts, 2005; Nosofsky, 1997) assume the existence of stable long term representations of the stimuli that are accessed on each trial and compared to the current stimulus. Although assuming a greater number of representations, it is well established that long term memory has a huge capacity and can be readily accessed. Suggesting that this resource is utilised may make more sense than relying on capacity limited short term memory.

The purpose of the present paper was to further investigate the representation processes utilised in absolute identification. The approach taken was to examine the viability of the relative judgement and exemplar comparison processes by asking participants to perform tasks that were based on these using a stimulus set from an absolute identification task. Performance on these tasks was then compared to the pattern of performance in an absolute identification task. Similar patterns of performance between tasks would indicate a common set of underlying processes, divergence, different underlying processes. In Experiment 1 participants completed an absolute identification of line length task (N = 8) and a simultaneous matching task in which all eight stimuli were presented in a random order

around a central target stimulus and participants were required to report the location (labelled 1-8) of the stimulus matching the target. The simultaneous matching task thus required comparison of the target stimulus with the stimulus set. If absolute identification involves a similar comparison process in which the target is compared to items retrieved from memory, then we would expect a similar pattern of findings between these two tasks. In Experiment 2, a single stimulus was presented on each trial (participants were initially shown all stimuli). The task was to indicate the size of the jump required to get from the previous stimulus on trial *n-1*, to the current stimulus on trial *n*. Thus participants were explicitly asked to perform part of the relative judgement process outlined by the RJM. If relative judgement is a major component underling absolute identification performance, then assessing the accuracy of jump estimates toward each stimulus should produce performance comparable to that of the absolute identification task.

Experiment 1

Participants. Participants were 22 university students who completed the experiment for course credit. All completed both the absolute identification task and the simultaneous matching task. The order in which these were attempted was counterbalanced. These participants also performed a sequential matching task (identical to the simultaneous matching except that the target was shown prior to the presentation of the stimulus array) that we do not report here. One participant was removed from analysis due to chance level responding in both tasks.

Materials and Stimuli. Participants sat approximately 100 cm from the centre of a 17 inch Sony CPD-G200 monitor used to display stimuli and instructions. The monitor used a

resolution of 1024 x 760. Responses were made using the numeric keys (1-8) on a standard USB keyboard. Stimuli consisted of eight white lines varying in length (along the horizontal plane) presented on a black background. The smallest stimulus was 70 pixels (2.1 cm) in length and 1 pixel in height with each subsequent stimulus increasing by 16% in length such that the largest stimulus was 128 pixels (5.9 cm) in length.

Design and Procedure.

Absolute identification. Participants were first shown each stimulus sequentially (from shortest to longest) along with its numeric label for 1,300 ms. After this participants then began the response phase of the task. On each trial, a small white fixation cross was presented centrally for 300 ms followed by a stimulus that remained onscreen until response. Stimuli were presented centrally with jitter of 50 pixels up or down and 50 pixels left or right (to prevent using cues such as the edge of the screen to aid judgements). The mapping of stimuli to keys was counterbalanced. For half of the participants 1 referred to the shortest stimulus and 8 the longest stimulus, with the reverse being true for the other participants. Following a response, the stimulus label was displayed for 750 ms. There was an inter-trial interval of 1,000 ms. There were 25 presentations of each stimulus with the order of presentation randomised across the whole session. Participants were given the opportunity for a short break every 100 trials.

Simultaneous matching. On each trial a small white fixation cross was presented centrally for 350 ms. The target stimulus was then presented at the centre of the screen. At the same time stimuli 1-8 were then presented in a circle (radius 300 pixels) surrounding the target in a randomized location (there were 8 possible locations), with each stimulus centred on the

circumference of the invisible circle and these centres being equally spaced along the circle.

Above each stimulus was a numeric label. Labels were associated with locations which were

labelled 1-8 in clockwise order. Participants were asked to find the stimulus on the outer

circle that matched the target stimulus presented centrally and respond with the appropriate

location. After response the display offset and then participants were provided with the

correct location, presented centrally for 400 ms. The inter trial interval was 600 ms. Each

target was shown 25 times, with the order in which targets were presented randomised across

the whole session. Participants were given the opportunity for a short break every 50 trials. In

both the absolute identification and simultaneous matching task participants were instructed

to respond as quickly and as accurately as possible.

Results and Discussion

In the absolute identification task responses were excluded if the response time was longer

than 3,000 ms or, for each participant, if their response time was greater than two standard

deviations above their mean RT. This led to exclusion of 8.5% of responses. In the

simultaneous matching task responses were excluded if the response time was longer than

6,000 ms or, for each participant, if their response time was greater than two standard

deviations above their mean RT. This led to exclusion of 8.3% of responses. The cut off

timings were felt adequate time to have completed the task. There was no evidence of an

effect of the order in which participants completed the tasks, and so we do not include it as a

factor in the following analyses.

Figure 1A shows the proportion of correct responses and mean RTs for the absolute

identification task. Bow effects are evident in both accuracy and RT, with stimuli toward the

ends of the range responded to faster and more accurately. A repeated measures ANOVA showed a significant effect of stimulus on proportion correct ($F(7, 140) = 9.14$, $\eta_P^2 = .31$, p <.001) and on RT ($F(7, 140) = 2.2$, $\eta_P^2 = .10$, p =.036). Stimulus/response confusions (Figure 1B) show that participants were clearly able to discriminate between stimuli.

Figure 1C also shows proportion of correct responses and mean RTs for the simultaneous matching task. A repeated measures ANOVA showed a significant effect of stimulus on proportion correct ($F(7, 140) = 9.59$, $\eta_P^2 = .32$, p <.001) and on RT ($F(7, 140) = 13.18$, $\eta_P^2 = .40$, p <.001). Stimulus/response confusions show that participants were clearly able to discriminate between adjacent stimuli (Figure 1D) As can be seen, the qualitative pattern of data matched that of the absolute identification task for both mean responses (panels A and C) and choice proportions (panels B and D). Clear bow effects were observed both in RT and proportion correct, with slower and less correct responding for stimuli in the middle of the range. However, performance did differ to that of absolute identification in some respects. Responding was more accurate in the matching task, probably because simultaneously presenting all the stimuli eliminates the need to access representations in memory. Nevertheless, responses were still far from perfect, and showed a response bow which was of a similar magnitude to that in the identification task. This suggests a similar underlying cause of this performance limitation, i.e., confusability between display stimuli in the matching task and items in memory in the identification task. That said, the bow was a bit flatter in the range from stimulus 2-7 for the matching compared with absolute identification. In terms of RTs, these were considerably slower in the matching task, and the RT bow was of a greater magnitude. This is probably because the matching task included a process in which attention would have been directed to display items that were similar to the target. This sequential

checking component of the task would slow down responding and would have greater effect for more central stimuli that have, on average, greater similarity to the whole set of stimuli in the display.

To examine whether the bow effects between tasks differed significantly a mixed ANOVA was run on proportion correct and RT data with stimulus (1-8) as a within subjects factor and task (absolute identification or simultaneous matching) and as a between subjects factor. The result of interest was the interaction between task and stimulus, which was significant in the ANOVA on proportion correct ($F(7, 140) = 2.93$, $\eta_P^2 = .13$, $p = .007$) and in the ANOVA on RT ($F(7, 140) = 7.79$, $\eta_P^2 = .28$, $p < .001$). These significant interactions suggest important differences in the shapes of the observed bows. For example, the bow in the matching task is deeper than the bow in the absolute identification task. Thus, whilst the broad similarity in the data in both tasks implies some common underlying processes, the significantly different bows suggest the presence of other, perhaps distinct, processes contributing to performance in these tasks. Nonetheless, the overall impression is that the data show similar qualitative patterns.

Experiment 2

Participants. Participants were 16 university students who either received course credit or were paid £5 (approximately $8) to complete the task.

Materials and Stimuli. The same materials and stimuli as Experiment 1 were used, with the exception that participants used the keys 0-7 to respond and a CTX EX951F 19 inch monitor

was used to display stimuli and instructions at a resolution of 1024 x 760. Stimuli ranged

from 70 to 128 pixels (2.4 - 6.8cm) in 16% steps up from the smallest stimulus.

Design and Procedure. The experiment was identical to the absolute identification task in

Experiment 1 with the following exceptions. Participants were told that they would see eight

lines of different lengths and that their task was to judge how different the stimulus was on

the current trial to the stimulus seen on the previous trial. They were told that, if the line on

the current trial is the line that was the next size up or next size down from the line on the

previous trial to press 1. If the line on the current trial is two sizes (jumps) up or down from

the line on the previous trial to press 2 etc. They were told the largest jump size was 7. Before

the response stage participants were shown all eight stimuli, but without their numeric labels.

Feedback given was the actual jump size. On the first trial participants were told to simply

make a random response as there was no trial preceding this to make a relative judgement.

Results

The first trial from each participant was discarded from analysis (as no relative judgement

was possible). Responses were excluded if the RT was longer than 3,000 ms or, for each

participant, if their response time was greater than two standard deviations above their mean

RT. This led to the exclusion of 10% of responses.

Response accuracy and RT for each of the jump sizes are shown in Figure 2A. A repeated

measures ANOVA showed a significant effect of jump size on proportion correct ($F(7, 105)$

$= 8.56$, $\eta_P^2 = .37$, $p <.001$) and on RT ($F(7, 105) = 11.53$, $\eta_P^2 = .43$, $p <.001$). There are

several striking results. First is that the task was hard, with poor accuracy. Chance responding

was at 12.5% meaning that larger jump sizes were not correctly identified at above chance

levels. Even when the stimulus was exactly the same as preceding stimulus, response

accuracy was only at 35%. This suggests that participants found it very difficult to maintain

stimulus representations from the previous trial. Second, with the exception of jump size 1,

there appeared a trend toward less accurate responding as jump size increased. Some light

can be cast on this pattern by examining confusability between the response and the correct

response (Figure 2B, percentage responses for each jump size are also shown in Figure 2C).

Whilst there were many more frequent small jump size responses (small jump sizes were

more common) participants were able to differentiate between the different jump sizes to

some extent. Thus a response of 2 was the most common response to jump size 2, but it was

also confused more with more similar jump sizes (e.g., 1 and 3). Nevertheless, the ability to

differentiate between different jump sizes was nowhere as clear as the ability to differentiate

between different stimuli in the absolute identification task (Experiment 1). Indeed, small

jump size responses were still given for larger jump sizes (e.g., >3) which reflect quite large

errors, indicating again that participants memory for the previous stimulus was relatively

poor. There was a clear pattern in the RT, with longer responding the larger the jump size.

Examining the jump size accuracy data, it can be assumed that, if relative judgement is used

in absolute identification, a correct estimate of jump size would translate into correct

identification of the stimulus if this were an absolute identification task. As such, for each

trial we added the jump size response to the previous stimulus in order to generate a predicted

absolute identification response for that trial. Note that this analysis assumes that participants

had access to the correct stimulus information on the previous trial. Although doubtful in this

task, in absolute identification feedback is typically provided, and so this analysis is a useful

means of assessing whether the ability to complete relative judgement results in data patterns

that would be expected in absolute identification. Figure 2D therefore plots predicted

response accuracy and RT as a function of stimulus. A repeated measures ANOVA on this

data showed no significant effect of stimulus on proportion correct ($F(7, 105) = 1.73$, η_P^2

$= .10$, $p = .11$) but a significant effect of stimulus on RT ($F(7, 105) = 5.19$, $\eta_P^2 = .26$, $p < .001$).

The predictions are surprising. First, response accuracy is low, much lower than observed in

the absolute identification task (Experiment 1). Secondly, although no significant effect of

stimulus was observed, if anything responses appear to be *more* accurate in the centre of the

stimulus range. Response times were predicted to be slowest for stimulus eight with some

evidence for slower responses toward the ends of the stimulus range in comparison to more

central stimuli. This may be the result of small jump sizes being faster and more accurate. If

shown a stimulus in the centre of the range, the size of the jump is much more constrained

(e.g., for stimulus 4 the maximum jump to it would be 3, whereas for stimulus 8 the

maximum jump would be 7).

In a similar vein, to the above analysis the absolute identification data can be re-plotted with

proportion correct as a function of the jump size between stimulus *n* and stimulus *n-1* (Figure

3). This shows bow effects in both accuracy and RT and a repeated measures ANOVA

showed a significant effect of stimulus on proportion correct ($F(7, 147) = 7.19$, $\eta_P^2 = .26$, p

$< .001$) and RT ($F(7, 147) = 22.07$, $\eta_P^2 = .51$, $p < .001$). Notably comparison between Figure 3

and Figure2A shows clear differences between data from the absolute identification task and

the relative judgment task.

Overall, the relative judgment data reveal a number of important findings. First, relative

judgement is not easy, it is difficult and prone to large errors, even when the stimulus is

repeated from the previous trial. Second, relative judgement takes longer and might be more difficult the larger the jump size. Third, predicting absolute identification success from success at relative judgement reveals a strikingly different pattern than that observed in absolute identification: standard bow effects are not produced and may even be reversed. Similarly, predicting relative judgement success from absolute identification success shows clear differences in the pattern of data to that found in the relative judgement task

General Discussion

In two experiments, we examined the evidence for relative judgement versus an exemplar comparison process in absolute identification. The approach taken was to assess performance in different tasks that required use of the processes similar to those hypothesised by the theories, with similar findings across tasks being taken as evidence for similar underlying processes. The simultaneous matching task (Experiment 1) required comparison of the target stimulus to display stimuli (a similar comparison process is hypothesised to take place when comparing a stimulus to exemplars in memory) and showed data patterns similar to the absolute identification task (Experiment 1) including bow effects in RT and accuracy and stimulus-response confusions. In contrast, the relative judgement task (Experiment 2) highlighted that relative judgement is difficult, error prone and that increasing jump size increases the time required to make a relative judgement without influencing accuracy. When assessing performance as a function of jump size in absolute identification, these latter trends are not observed. Furthermore, plotting accuracy as a function of stimulus in the relative judgement task shows that relative judgement, if required to correctly identify a stimulus, would produce bow effects in accuracy and RT in the opposite direction to that observed in absolute identification. Overall, the data indicate that absolute identification is not likely to

rely on a relative judgement process that uses jump size information, such as that proposed by Stewart et al. (2005).

Caveats are required, however. The relative judgement task we used is not a reflection of the exact process assumed by the RJM (Stewart et al., 2005). In the relative judgement task the judgement is explicit, but it is not clear in the RJM whether the relative judgement process in absolute identification is explicit or implicit. If implicit, then it could be argued that the relative judgement task does not directly utilise this process. However, such an argument would require an explanation of why implicit knowledge about previous stimuli and/or jump size is available for use in absolute identification but not in the relative judgement task. Another difference is that the RJM assumes the use of distance estimates from multiple previous trials, whereas the task here did not take into account previous distances. Similarly, the RJM can take into account feedback. In absolute identification, feedback provides implicit information about jump size. For example, if the stimulus shown on trial *n-1* was a 4, then the range of jump sizes on trial n would be 0-4. In comparison, because feedback in the relative judgement task was about the size of the jump (and not the location within the set of the stimuli) this information could not be used in the same way. Although this difference might explain some of the errors in the relative judgement task, it still does not offer a full explanation as to why performance was so poor when a stimulus was repeated on successive trials (when performance in absolute identification for repeat stimuli is typically very high). A more parsimonious explanation is that participants simply found it difficult to retain and utilise information from the previous trial. Some of this difficulty may be due to the use of a logarithmic scale to generate stimuli. This may increase task difficulty as the same jump size would differ in terms of physical magnitude depending on where it was on the stimulus range (in particular from a given point the same jump size up or down a scale would differ in

physical magnitude). Nevertheless, the RJM is based on the notion that stimulus scales are logarithmic, and calculation of a difference between the current and preceding stimulus in the RJM is given by $ln(Xn/Xn\text{-}1)$ in which X refers to the physical magnitude of the stimulus. Thus the experiment offers a good test of the ability to make relative judgements this way. As such, we believe that the difficulty of the relative judgement process means that it is unlikely that relative judgement is a central process in absolute identification.

For exemplar-based theories (Guest, Kent & Lamberts, 2015, Kent & Lamberts, 2005, Nosofsky, 1997) the data are more promising: the similarity between the data in the absolute identification task and the simultaneous matching task suggest that these two tasks might involve similar underlying processes. As the matching task explicitly involves comparison between the target and display stimuli, the suggestion is that the absolute identification task might also rely on a comparison process, but between the stimulus and exemplars in memory. Caution must be taken with this interpretation as data are correlational, and visual comparison may involve different processes than memory-based comparison. Nevertheless, given that relative judgement leads to such poor performance, it seems inevitable that some long term referent must be accessed during absolute identification. Limited capacity models suggest that this is information about the experimental context (particularly the end anchors). However, in the simultaneous matching task a process based on the use of the end stimuli alone would not be very helpful. Given that the simultaneous matching data and the absolute identification data are very similar, it suggests the whole range of stimuli are represented, at least to some extent in absolute identification. Moreover, the most parsimonious explanation of the data is that a similar comparison process underlies both tasks. Indeed, one of the important aspects of the exemplar models of absolute identification is that they have also provided a very good account of data from other tasks including categorisation (Nosofsky,

1996; Lamberts, 2000, Guest & Lamberts, 2010), recognition (Brockdorff & Lamberts, 2000; Nosofsky, 1997; Zaki & Nosofsky, 2001), same-different judgements (Cohen & Nosofsky, 2000), and visual search (Guest & Lamberts, 2011). This suggests that these processes are fundamental to cognition, adding weight to the notion that the similarity of the pattern of responses in absolute identification and simultaneous matching is an indication of a common underlying process.

A major shortcoming of exemplar models of absolute identification is that they have not been extended to account for sequential effects. One way for them to capture sequential effects is to adopt a similar approach to SAMBA in which they are produced by decay. Each response has an associated response counter which races toward a response threshold. More recent stimuli receive greater activation based on the premise that these stimuli receive selective attention. Indeed, exemplar models already have a parameter that represents memory strength and is affected by factors such as recency (Equation 3, Nosofsky & Palmeri, 1997, for the EBRW and Equation 12, Lamberts, 2000, for the EGCM-RT; see also Kent & Lamberts, in press). Although this mechanism may help capture assimilation and repetition advantages, it does not obviously allow the models to predict contrast. A further issue is that assimilation does not appear to be influenced by the time between trials (Donkin et al, 2015; Matthews & Stewart, 2009), whereas the decay mechanism in SAMBA predicts decreasing assimilation due to activation for response counters decaying and decreasing to baseline during this time (Donkin et al. 2015). It may be that decay is partially interference-based rather than time-based. Nevertheless, the fact that sequential effects can be produced without recourse to relative judgement is important in light of the present results which suggest that relative judgement is surprisingly difficult.

The data presented here are consistent with absolute identification being subserved by an exemplar comparison process and not by a relative judgement process. The difficulty of the relative judgement task either indicates that calculations of distance are difficult, or that it places large demands on the resources of a limited capacity short term memory. In absolute identification performance is improved if the stimulus is the same or very similar to the preceding stimulus and this has been taken as evidence of a relative process (Brown et al., 2008). Interestingly, we do show that relative judgement is easier when the jump size is small and so it may be that a relative judgement process is utilised when jump size is small. Indeed, Rouder et al. (2004) suggested that practice may improve absolute identification primarily because people get better at computing relative differences when the jump size is large. However, Dodds et al. (2011) found that practice reduces contrast but not assimilation, which may not be consistent with relative judgement increasing with practice. Interestingly, in order to capture the extent of increased accuracy when stimuli were repeated, Brown et al. (2008) had to integrate a relative judgement component into SAMBA by assuming that, on some proportion of trials, one of the end anchors was replaced with the previously seen stimulus. Although this version of the model provided a good fit of the repeated stimulus effect observed in Lacouture (1997) it assumed that relative judgement was used on 62.5% of trials. Given the difficulty participants had with making explicit relative judgements in our Experiment 2 we believe that this might considerably overestimate the extent of relative judgement. Rather, we suspect that relative judgement is used only on a small number of trials, if at all. It remains to be seen the circumstances under which participants switch from reliance on long term referents to a relative judgement (or vice versa)

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods, 44*, 158-175

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(1), 77–102. http://doi.org/10.1037/0278-7393.26.1.77

Brown, S.D., Marley, A.A.J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review, 115,* 396-425.

Brown, S. D., Marley, A. A. J., Dodds, P., & Heathcote, A. (2009). Purely relative models cannot provide a general account of absolute identification. *Psychonomic Bulletin & Review*, *16*(3), 583–593. http://doi.org/10.3758/PBR.16.3.583

Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same--different judgments. *Journal of Experimental Psychology. Human Perception and Performance*, *26*, 1549–1569. http://doi.org/10.1037/0096-1523.26.5.1549

Cohn, J., (1900) Gefühlston und Sattigung der Farben, *Philos. Stud*., *15,* 279-286

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185

Dodds, P., Donkin, C., Brown., S. D., & Heathcote, A. (2011). Increasing capacity: Practice effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 477-492.

Dodds, P., Donkin, C., Brown, S. D., Heathcote, A., & Marley, a a J. (2011). Stimulus-specific learning: disrupting the bow effect in absolute identification. *Attention, Perception & Psychophysics*, *73*, 1977–1986. http://doi.org/10.3758/s13414-011-0156-0

Donkin, C., Brown, S.D., Heathcote, A. & Marley, A.A.J. (2009). Dissociating speed and accuracy in absolute identification: The effect of unequal stimulus spacing. *Psychological Research, 73*, 308-316

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the Linear Ballistic Accumulator model. *Journal of Mathematical Psychology*, 55, 140-151.

Donkin, C., Chan, V., & Tran, S. (2015). The effect of blocking inter-trial interval on sequential effects in absolute identification. *Quarterly Journal of Experimental Psychology, 68*, 129-143.

Donkin, C., Rae, B., Heathcote, A., & Brown, S. D. (in press). Why is labelling simple magnitudes so hard? *Oxford Handbook of Computational and Mathematical Psychology*.

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

Guest, D., Kent, C., & Adelman, J. S. (2010). Why Additional Presentations Help Identify a Stimulus. *Journal of Experimental Psychology-Human Perception and Performance, 36*, 1609-1630.

Guest, D., Kent, C., & Adelman, J. S. (in submission). The relative importance of perceptual and memory sampling processes in determining the time course of absolute identification

Guest, D., & Lamberts, K. (2010). The prioritization of perceptual processing in categorization. *Attention Perception & Psychophysics, 72*(4), 1079-1096.

Guest, D., & Lamberts, K. (2011). The time course of similarity effects in visual search. *Journal of Experimental Psychology-Human Perception and Performance, 37*, 1667-1688.

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning Memory and Cognition, 31,* 289-305.

Kent, C., & Lamberts, K. (in press). Stimulus Probability Effects in Absolute Identification. *Journal of Experimental Psychology: Learning Memory and Cognition,*

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.

Lacouture, Y., & Marley, A.A.J. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics, 66*, 1206-1226.

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

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

Lockhead, G. R., & Hinson, J. (1986). Range and sequence effects in judgment. *Perception & Psychophysics,* 40, 53–61.

Lockhead, G. R. (2004). Absolute judgments are relative: A reinterpretation of some psychophysical ideas. *Review of General Psychology, 8,* 265–272.

Matthews, W. J., & Stewart, N. (2008). The effect of stimulus range on two-interval frequency discrimination. *The Journal of the Acoustical Society of America*, *123*(April), EL45–L51. http://doi.org/10.1121/1.2884084

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.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychological judgments applied to magnitude estimation. *Journal of Mathematical Psychology, 30,* 339–390.

Miller, G.A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

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

Nosofsky, R.M. (1997). An exemplar based random-walk model of speeded categorization and absolute judgement. In A. Marley, A.J. (Ed.), *Choice, Decision, and Measurement* (pp. 347-365). NJ: Lawrence Erlbaum Associate.

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

Petrov, A.A., & Anderson, J.R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review, 112*, 383-416.

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

Rouder, J. N., Morey, R. D., Cowan, N., & Pfaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review, 11,* 938–944.

Stewart, N., Brown, G.D.A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*, 881-911.

Stewart, N., & Matthews, W. J. (2009). Relative judgment and knowledge of the category structure. *Psychonomic Bulletin & Review*, *16*(3), 594–599. http://doi.org/10.3758/PBR.16.3.594

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

Wever, E. G., & Zener, K. E. (1928). The method of absolute judgment in psychophysics. *Psychological Review, 35,* 466–493.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1022-1041.
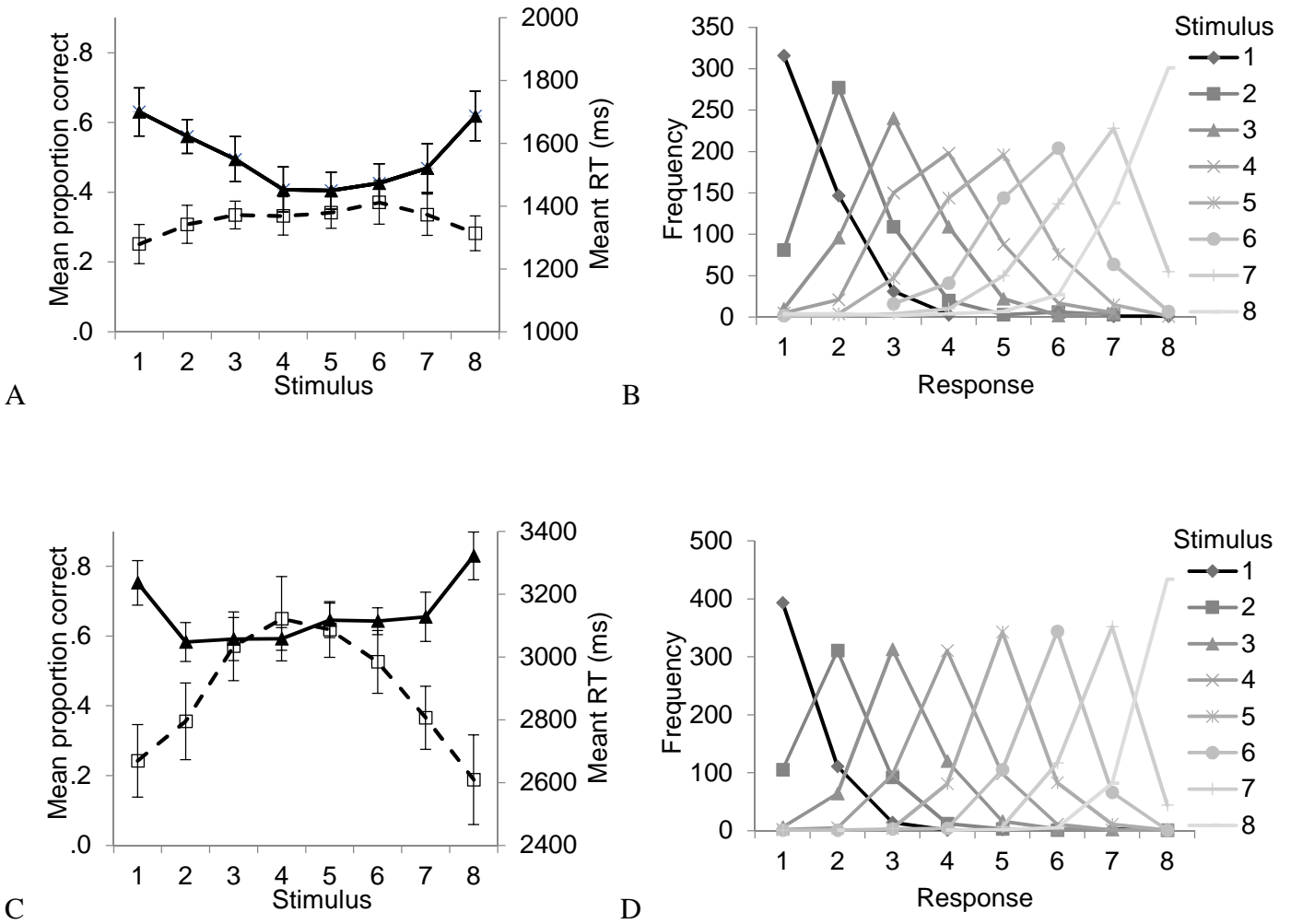
Figure List

Figure 1. Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each stimulus in the absolute identification task (A) and simultaneous matching task (C) in Experiment 1. Error bars represent 95% confidence intervals (see Baguley, 2012). Panels B and D show stimulus response confusion matrices for the identification task and matching task respectively.

Figure 2. Data from the relative judgment task in Experiment 2. A) Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each jump size B) Frequency of responses for each jump size C) Percentage of responses made for each response at each jump size D) Predicted identification accuracy (filled line, left axis) and RT (dashed line, right axis) for each stimulus. Error bars represent 95% confidence intervals.

Figure 3. Data from the absolute identification task in Experiment 1. Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each jump size

Figure 1. Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each stimulus in the absolute identification task (A) and simultaneous matching task (C) in Experiment 1. Error bars represent 95% confidence intervals (see Baguley, 2012). Panels B and D show stimulus response confusion matrices for the identification task and matching task respectively.
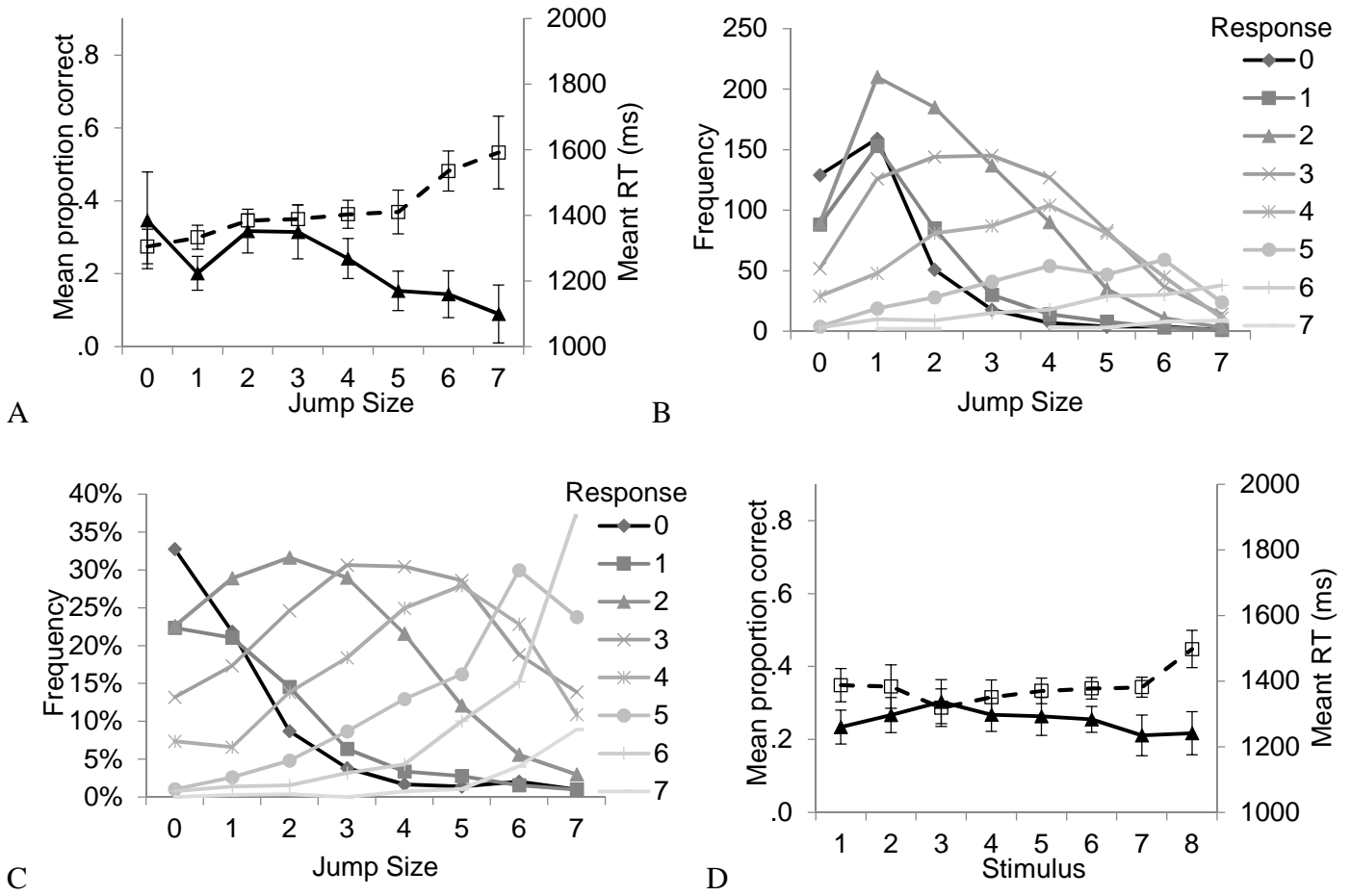
Figure 2. Data from the relative judgment task in Experiment 2. A) Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each jump size B) Frequency of responses for each jump size C) Percentage of responses made for each response at each jump size D) Predicted identification accuracy (filled line, left axis) and RT (dashed line, right axis) for each stimulus. Error bars represent 95% confidence intervals.
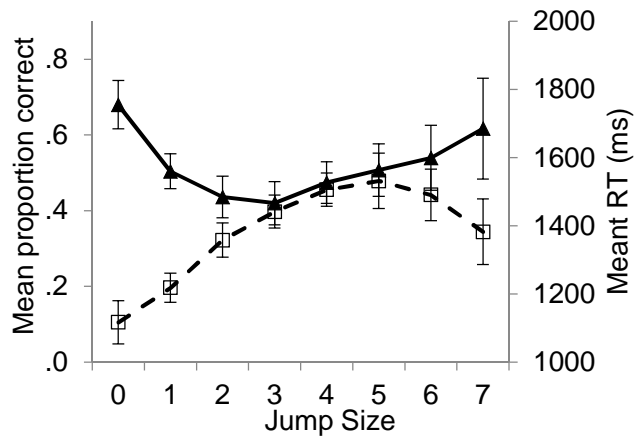
Figure 3. Data from the absolute identification task in Experiment 1. Proportion correct (filled line, left axis) and RT (dashed line, right axis) for each jump size