

# When and Why is $9 > 221$ ? Reference Sets Evoked by Elicitation Methods and Stimuli

Lim M. Leong<sup>1</sup>, Craig R. M. McKenzie<sup>1,2</sup>, Johannes Müller-Trede<sup>2</sup>, & Shlomi Sher<sup>3</sup>

<sup>1</sup>Psychology Department, UCSD, <sup>2</sup>Rady School of Management, UCSD, <sup>3</sup>Pomona College

## Abstract

Previous research has demonstrated the surprising finding that, between-subjects, 9 is judged larger than 221. A speculative explanation held that each focal stimulus (9 or 221) evokes a distinct “reference set” (single-digit or triple-digit numbers) for evaluation. In a series of experiments, we show that the focal stimulus indeed affects the reference set, but so do other aspects of the experimental context, including the rating scale. Thus, 9 exceeds 221 when ratings are made on a 10-point scale, but not on a continuous scale or a 1000-point scale. Reference sets are joint products of the stimulus and the judgment scale.

## Introduction

- When presented with a single number without context and asked to make a subjective rating in a between-subjects design, participants judged the number 9 to be larger than 221 (Birnbau, 1999)
- Birnbau (1999) postulated that 9 was more likely to suggest a reference set of single-digit numbers while 221 was more likely to suggest a reference set of triple-digit numbers
- Research on formal features of questionnaires suggests that the elicitation method may also influence the evoked reference set (e.g., Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991; McGraw, Larsen, Kahneman, & Schkade, 2010)

## Elicitation Methods

### A. Original Elicitation Method (used in Exp. 1 & 2)

On a scale of 1 to 10, where  
1 = very very small  
10 = very very large

Please judge, how large is the number 221?

### C. 1000-point Rating Scale (used in Exp. 2)

On a scale of 1 to 1000, where  
1 = very very small  
1000 = very very large

Please judge, how large is the number 9?

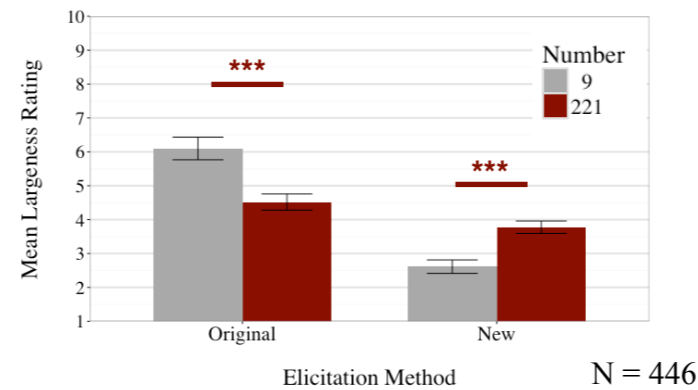
### B. Visual Slider Rating Scale Without Numerical References (used in Exp. 1, 3a, & 3b)

Please judge, how large is the number 221?

very very small  very very large

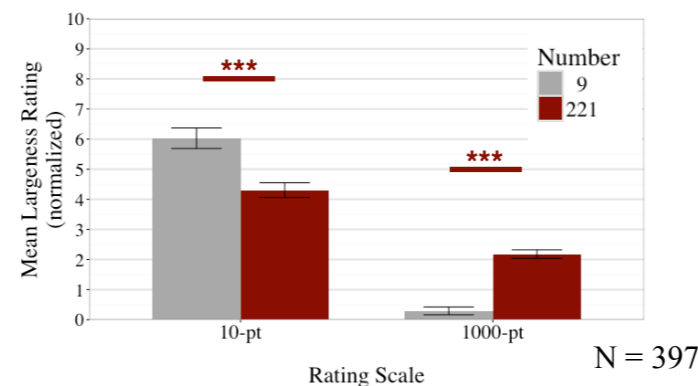
## Experiment 1

- The  $9 > 221$  effect depends crucially on how the ratings are elicited
- Participants rated the same numbers in Birnbau’s (1999) study using either the original elicitation method (A) or the new visual slider rating scale (B)



## Experiment 2

- The range is one feature of the elicitation method that affects the evoked reference set
- The 1 to 10 rating scale in the original elicitation method (A) is the same reference set that is thought to be evoked by the number 9
- The verbal anchors remain the same, but 9 and 221 are now both within the range of the new 1000-point rating scale (C)

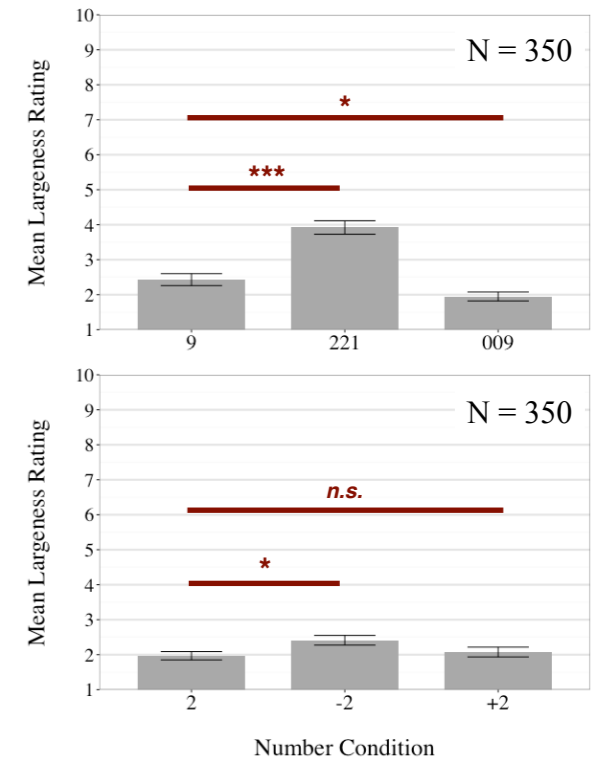


### References

Birnbau, M. H. (1999). How to show that  $9 > 221$ : Collect judgments in a between-subjects design. *Psychological Methods*, 4(3), 243-249.  
 McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science*, 21(10), 1438-1445.  
 Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, 49(3), 388-395.

## Experiment 3a & 3b

- We provide novel evidence for stimulus-evoked reference sets while using the same slider scale (B) as Exp. 1



## Conclusions

- We replicated Birnbau’s (1999)  $9 > 221$  effect when using his exact methods, but not in conceptual replications using a continuous slider scale or a 1000-point rating scale
- The elicitation method contributes to the reference set that is evoked
- Furthermore, we found new evidence that the focal stimulus matters, showing that  $9 > 009$  and  $-2 > 2$  on a slider scale
- Our data show that when the context is not specified, people construct a relevant context from features of the task environment rather than assuming there is no context