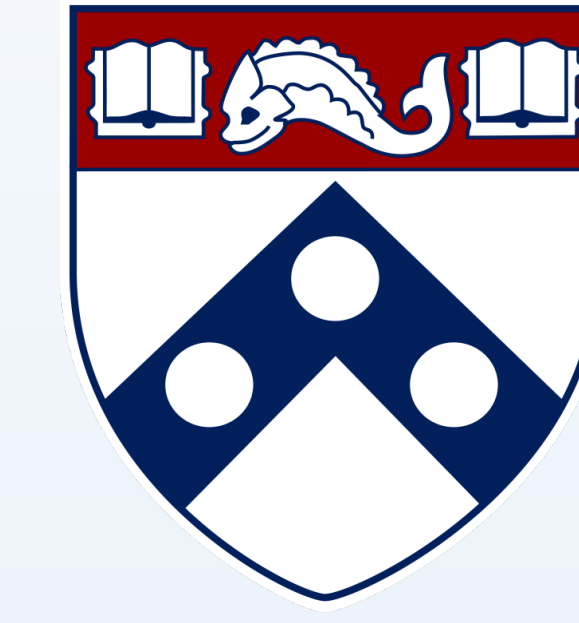


Memory Modeling of Counterfactual Retrieval



Feiyi Wang, Ada Aka, Sudeep Bhatia
University of Pennsylvania

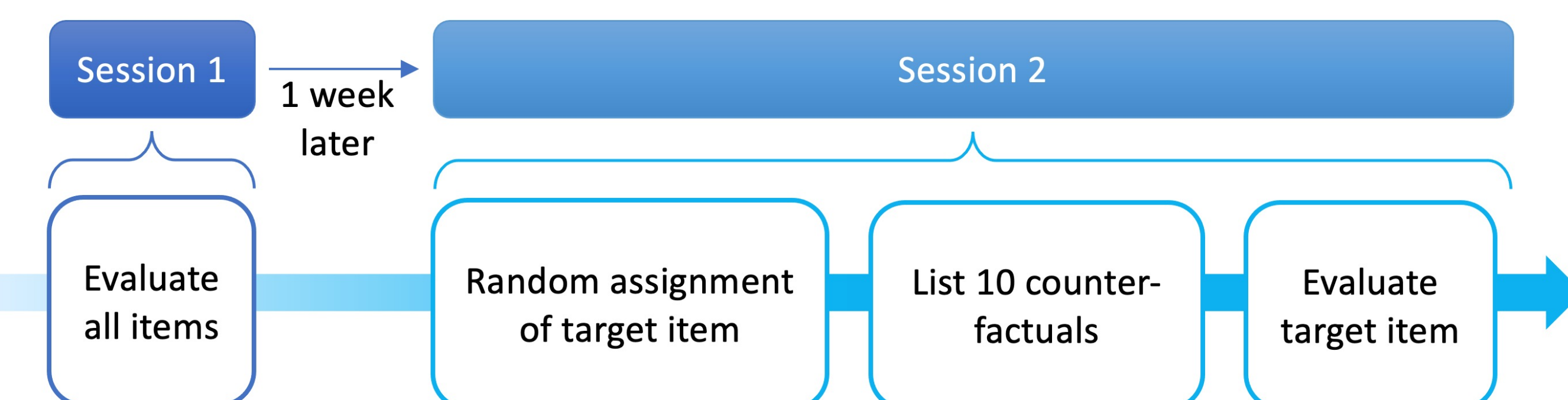
Introduction

- ◆ People often engage in counterfactual thinking (i.e., thinking about what's contrary to the fact—they imagine alternative possibilities to a target event).
- ◆ Whether a counterfactual thought comes to mind depends on its desirability (Phillips et al., 2019) and its semantic similarity with the target event (Kahneman & Miller, 1986).
- ◆ We build a formal quantitative model to examine the effects of previously suggested mechanisms as well as memory mechanisms on counterfactual retrieval.

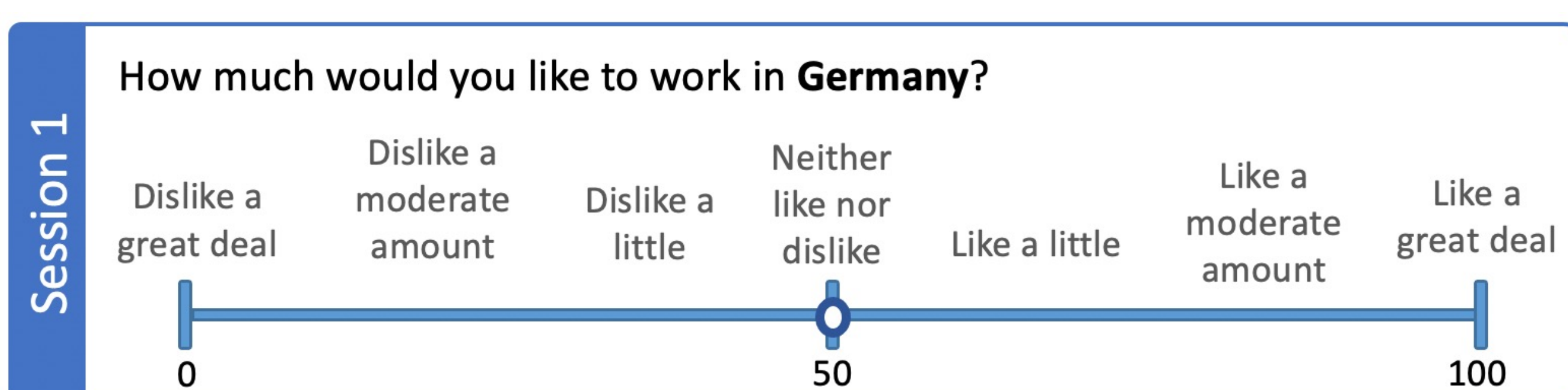
Methods

- ◆ Three online studies using similar procedures with different contexts:
 - Study 1 (job offer): $N = 53$; $M_{age} = 20$
 - Study 2 (vacation trip): $N = 53$; $M_{age} = 32$
 - Study 3 (fruits & vegetables): $N = 40$; $M_{age} = 20$

Schematic of experimental design



Example from Study 1 (target item is Germany):

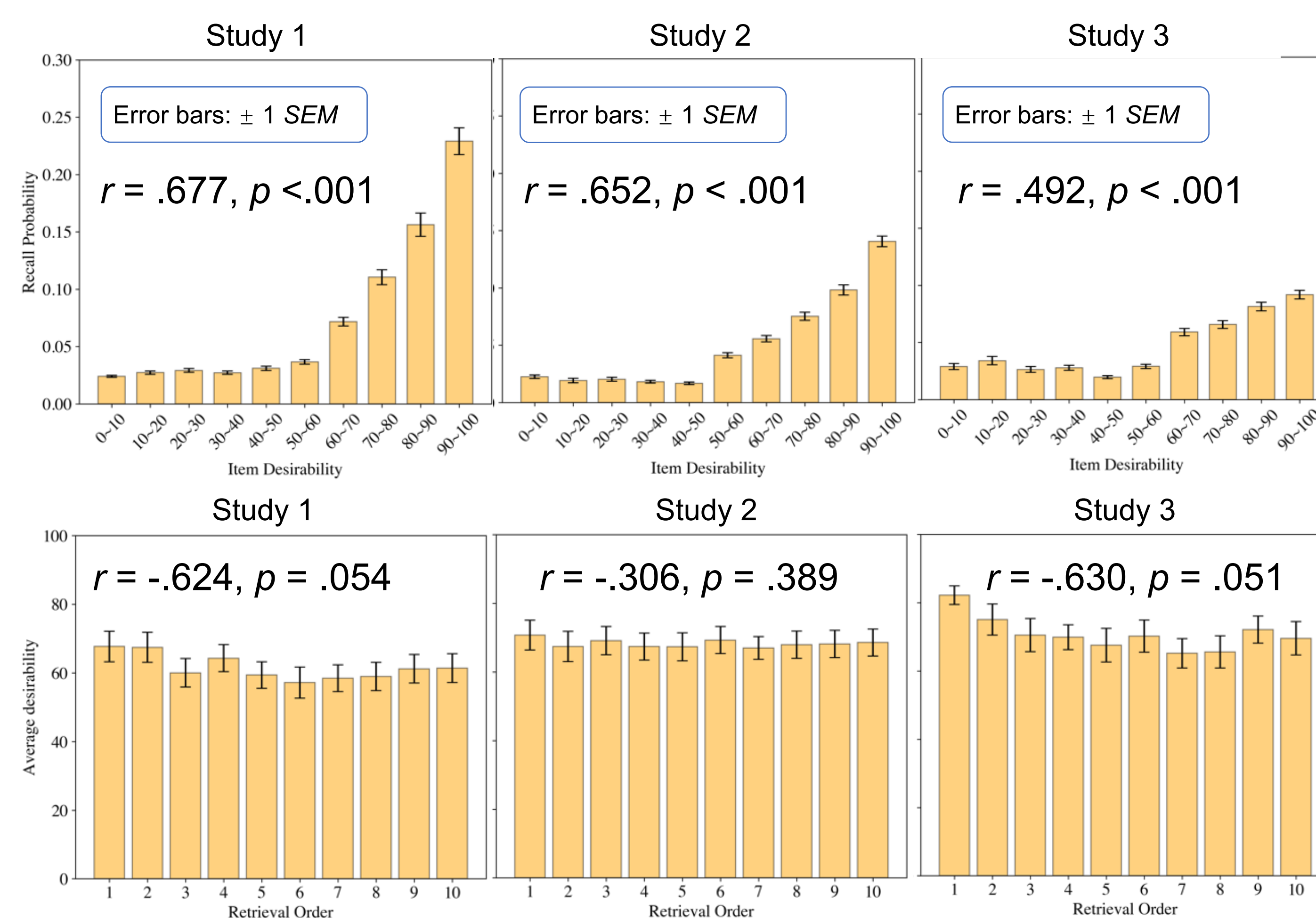


Session 2: Now, please list 10 other countries that come to your mind as you think about your job offer in Germany. Please list these countries in the order in which they come to your mind.

The Desirability Effect

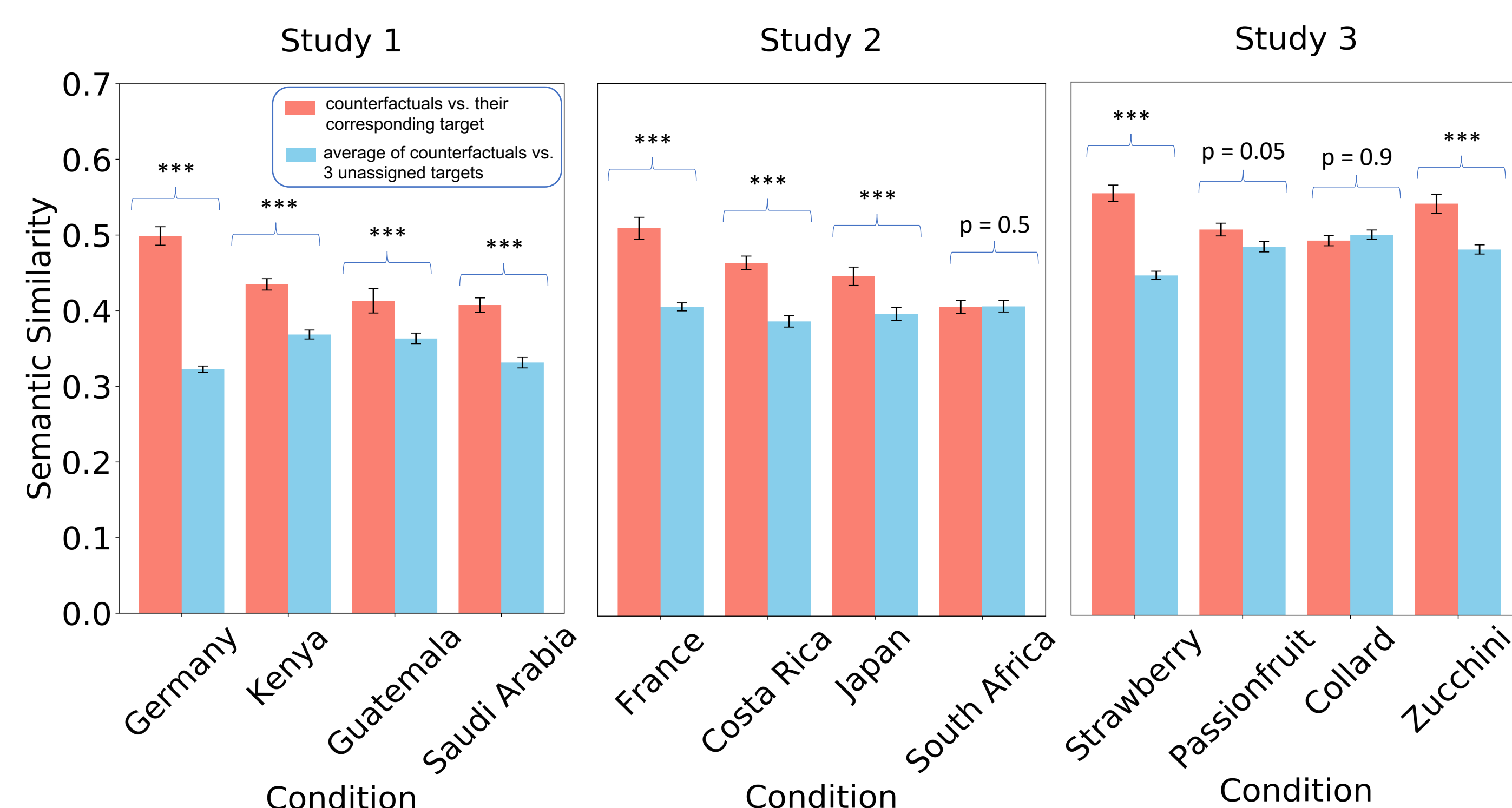
- ◆ A more desirable counterfactual thought is more likely to come to mind, earlier in the sequence.

$$P(\text{retrieval}) \text{ of an item } i = \frac{\text{total \# of times } i \text{ is retrieved}}{\text{total \# of participants}}$$



The Semantic Similarity Effect

- ◆ The semantic similarity between any two words is calculated by their cosine similarity ($\cos\theta$) in an English word vector model using Google's corpus.
- ◆ People tend to think about counterfactuals that are more semantically similar with the assigned target.



Markov Model Results

- ◆ Our model is a Markov random walk on a semantic network.
- ◆ We found that the desirability and the semantic similarity effects hold even after controlling for likelihood and other memory mechanisms.

$$\text{Activation}(\text{Spain}|\text{Cuba}, \text{Target}) = \beta_D \cdot \text{Desirability}(\text{Spain}) + \beta_L \cdot \text{Likelihood}(\text{Spain}) + \beta_{ST} \cdot \text{Sim}(\text{Spain}, \text{Target}) + \beta_{SP} \cdot \text{Sim}(\text{Spain}, \text{Cuba}) + \beta_F \cdot \text{Frequency}(\text{Spain})$$

	Study 1	Study 2	Study 3
Model	Coeff.	Coeff.	Coeff.
Desirability	0.282***	0.373***	0.184***
Likelihood	0.023	0.095	0.014
Sim. w/ target	1.896***	1.341***	2.118***
Sim w/ prev. item	6.773***	6.751***	6.105***
Word frequency	1.069***	0.923***	0.838***

*** $p < .001$, likelihood ratio tests between full model and a 1-variable reduced model. Models were fit using maximum likelihood estimation.

Conclusion

- ◆ We build a formal parametric model to investigate the mental processes at play during counterfactual thinking.
- ◆ The retrieval of counterfactuals is influenced by subjective desirability and how semantically similar it is with the target item.
- ◆ We have run additional studies to explore of priming on counterfactual retrieval.

References

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, memory, and cognition*, 34(5), 1191.

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval?. *Journal of Memory and Language*, 46(1), 85-98.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, 93(2), 136.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in cognitive sciences*, 23(12), 1026-1040.