# Modeling the Retrieval of Counterfactual Thoughts

Feiyi Wang, Ada Aka, Sudeep Bhatia
Psychology Department, University of Pennsylvania

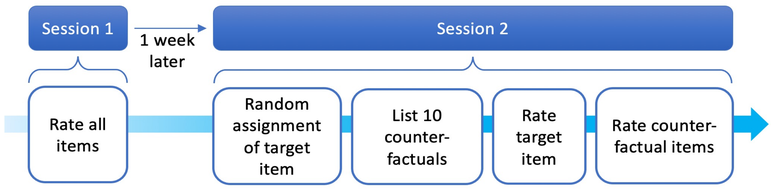
## Introduction

- People often engage in counterfactual thinking—they imagine how an event could have turned out differently (Kahneman & Miller, 1986).
- Whether a counterfactual thought comes to mind depends on its desirability (Phillips *et al.*, 2019), as well as memory mechanisms such as its semantic similarity with the actual event (De Brigard *et al.*, 2021).
- We build a Markov model to examine how counterfactual thoughts are retrieved from a set of all possible counterfactuals.

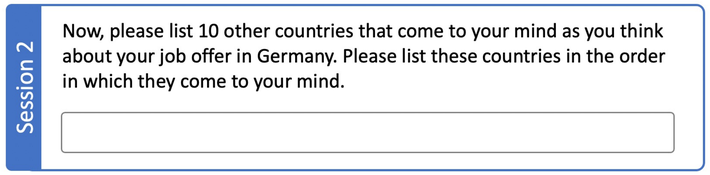
## Methods

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

**Schematic of experimental design**



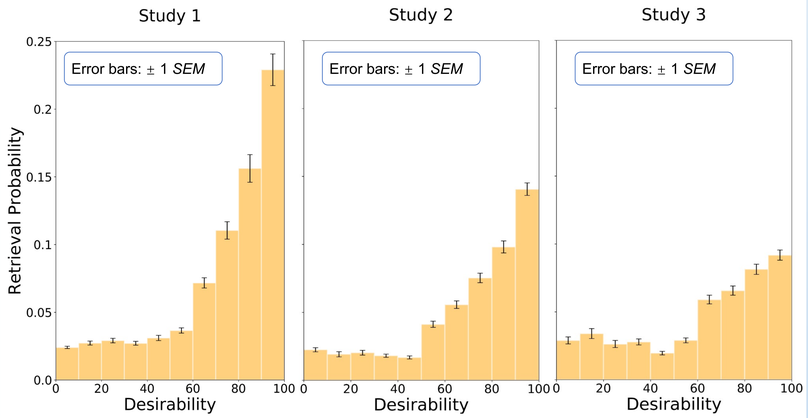

**Example from Study 1:**


## The Desirability Effect

- A more desirable counterfactual thought is more likely to come to mind.

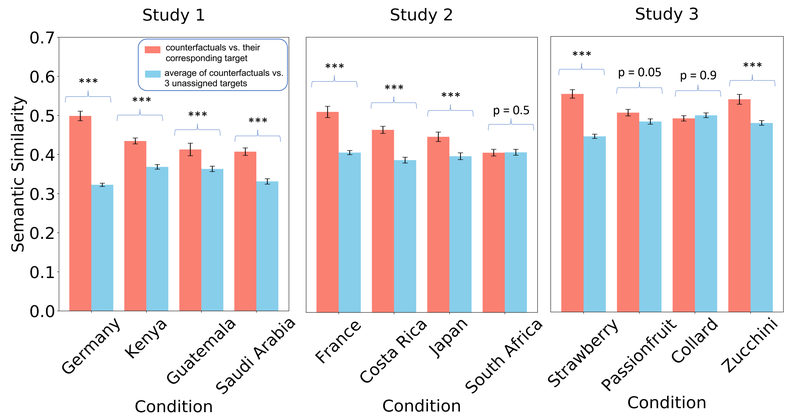
$$\text{P(retrieval) of an item } i = \frac{\text{total \# of times } i \text{ is retrieved}}{\text{total \# of participants}}$$




$r = .677$, $p < .001$ $r = .652$, $p < .001$ $r = .492$, $p < .001$

## 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 target item.




***$p < .001$

## Markov Model Results

- Full model variables: desirability, likelihood of occurrence, language frequency (logarithmized), semantic similarity with the previously retrieved item (see Howard & Kahana, 2002), and semantic similarity with the target item.
- Dropping desirability from the model significantly decreases the model's fit (i.e., it is a worse model).
- Dropping semantic similarity to the target from the model also significantly decreases fit.

| | Study 1 | Study 2 | Study 3 |
|---|---|---|---|
| **Model** | **-LL** | **-LL** | **-LL** |
| Full model | 2000.801 | 2076.384 | 3095.437 |
| w/o desirability | 2009.639*** | 2080.397* | 3107.395** |
| w/o similarity to target | 2016.790*** | 2085.610*** | 3108.972*** |

-LL: Negative Log Likelihood **$p<.01$. ***$p < .001$.

## Conclusion

- We build a formal parametric model to investigate the mental processes at play during counterfactual thinking.
- The retrieval of counterfactual thoughts is influenced by subjective desirability and how semantically similar it is with the reality.
- We have run additional studies to show the effect of desirability using experimental manipulation.

## References

De Brigard, F., Henne, P., & Stanley, M. L. (2021). Perceived similarity of imagined possible worlds affects judgments of counterfactual plausibility. *Cognition, 209*, 104574.
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.