Exploring Source of Variance in Decision Making with Large Language Models

Sudeep Bhatia

University of Pennsylvania

11/18/2023

SJDM 2023

Funding from the NSF

Psychometrics

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

APPENDIX A

For each of the following statements, please indicate the **likelihood** of engaging in each activity. Provide a rating from **1 to 5**, using the following scale:

1	2	3	4	5
Extremely unlikely		Not sure		Extremely likely

1. ________ Admitting that your tastes are different from those of your friends. (S)
2. ________ Arguing with a friend who has a very different opinion on an issue. (S)
3. ________ Asking your boss for a raise. (S)
4. ________ Betting a day's income at the horse races. (F)
5. ________ Buying an illegal drug for your own use. (E)

⋮

48. ________ Trying bungee jumping. (R)
49. ________ Using office supplies for your personal business. (E)
50. ________ Wearing unconventional clothes. (S)

Note: E = ethical, F = financial, H = health/safety, R = recreational, and S = social items.

Weber et al., 2002

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

	Risk behaviors ($N=547$)					
	Factor					
Item	1	2	3	4	5	6
Recreational						
6	0.14	0.02	0.01	**0.46**	0.10	−0.15
15	0.05	−0.02	0.02	0.27	0.05	0.29
18	−0.11	0.03	−0.09	**0.65**	−0.00	0.04
19	0.10	0.10	0.06	**0.43**	0.19	−0.09
20	−0.02	0.03	−0.06	**0.60**	−0.09	0.10
21	**0.38**	−0.05	−0.01	**0.34**	0.00	−0.01
22	−0.02	−0.05	0.04	**0.77**	−0.05	−0.07
26	−0.12	0.09	−0.11	**0.36**	0.11	0.21
39	−0.12	0.03	0.10	**0.74**	−0.09	0.02
48	−0.01	−0.03	0.09	**0.62**	−0.05	−0.04

	Risk behaviors ($N=547$)					
	Factor					
Item	1	2	3	4	5	6
Financial						
4	**0.39**	0.15	0.29	0.07	0.02	−0.01
9	**0.47**	0.05	−0.21	0.06	−0.03	0.02
29	0.06	**0.76**	−0.07	0.04	0.00	0.02
30	0.12	**0.67**	0.07	−0.05	−0.01	−0.04
31	−0.01	**0.70**	−0.02	0.03	−0.04	0.03
32	**0.45**	0.15	−0.01	0.05	0.09	−0.02
33	**0.53**	0.00	−0.02	0.10	−0.14	0.19
44	0.20	−0.08	**0.30**	−0.02	0.09	0.19
46	**0.33**	0.12	0.26	−0.01	0.17	0.07
47	0.15	0.10	0.01	0.00	0.15	0.04

Weber et al., 2002

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages

- Naturalistic items (e.g. common risks, behaviors, attitudes)

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages

- Naturalistic items (e.g. common risks, behaviors, attitudes)

Disadvantages

- Agnostic about underlying decision process
- Cannot make out-of-sample predictions

Decision Modeling

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

	Apartment 1	Apartment 2
Bedrooms	2	3
Size (sq ft)	1100	2000
Distance to city (mi)	0.5	4.9
Parking	X	✓

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

<u>Multi-attribute Utility Theory (MAUT):</u>

Option i's attributes values: $\mathbf{x}_i$

Person j's attributes weights: $\mathbf{w}_j$

Person j's utility for option i: $U_{ij} = \mathbf{w}_j \cdot \mathbf{x}_i$

Similarity between option i and i': $\mathrm{sim}(\mathbf{x}_i, \mathbf{x}_{i'})$

Similarity between person j and j': $\mathrm{sim}(\mathbf{w}_j, \mathbf{w}_{j'})$

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:

- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:

- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Disadvantages

- Artificial stimuli

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages
- Naturalistic items (e.g. common risks, behaviors, attitudes)

Disadvantages
- Agnostic about underlying decision process
- Cannot make out-of-sample predictions

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:
- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Disadvantages
- Artificial stimuli

How to synthesize?

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages
- Naturalistic items (e.g. common risks, behaviors, attitudes)

Disadvantages
- Agnostic about underlying decision process
- Cannot make out-of-sample predictions

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:
- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Disadvantages
- Artificial stimuli

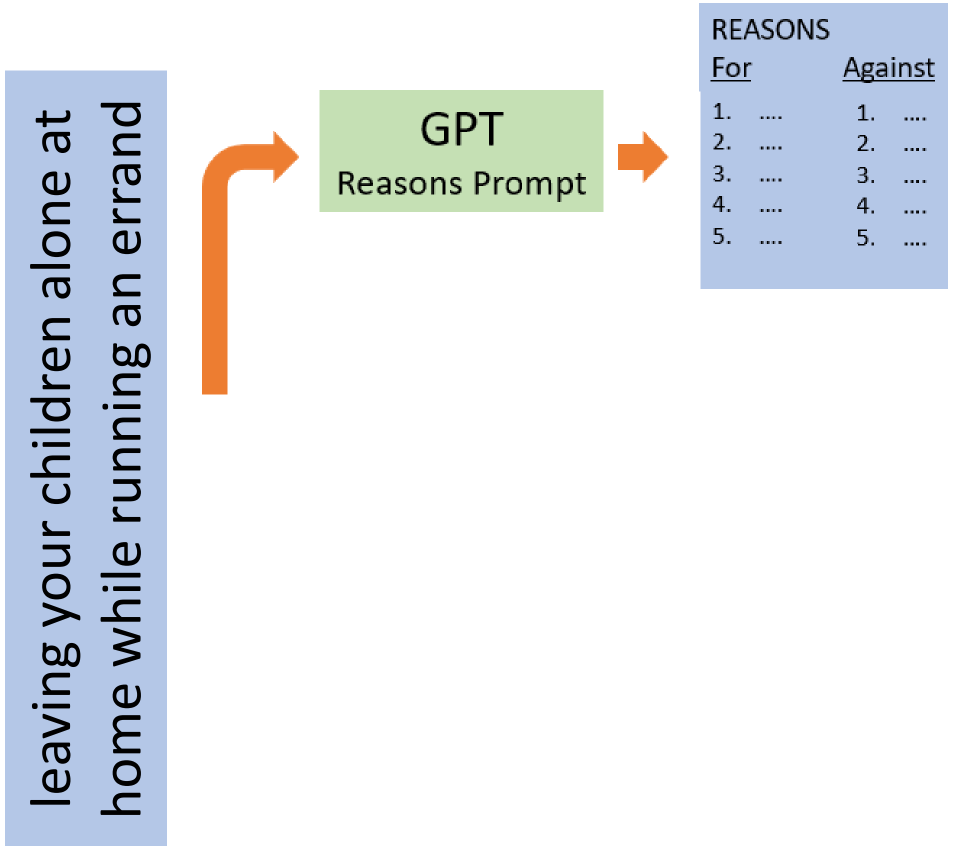
How to synthesize?

Answer: Large language models!

Overview of Approach

leaving your children alone at home while running an errand

Overview of Approach



Overview of Approach

leaving your children alone at home while running an errand

GPT
Reasons Prompt

REASONS

For	Against
1.	1. ....
2.	2. ....
3.	3. ....
4.	4. ....
5.	5. ....

Reasons for Leaving Young Children Alone at Home While Running an Errand:
#1 It is a good opportunity to give children a sense of **independence and responsibility**.
#2 Children can use it as an opportunity to study, exercise or practice a skill without interruption.
#3 It may help them **gain confidence** in their own abilities and trust in their parents.
#4 Errands can be done more efficiently without the need to supervise children's activities.
#5 It may be a **cheaper, more convenient** option than hiring a babysitter.

Reasons Against Leaving Young Children Alone at Home While Running an Errand:
#1 It is more difficult to monitor young children who are home alone and **ensure their safety**.
#2 It can be **emotionally and psychologically difficult** for some children who are not used to being alone.
#3 Some neighborhoods may not be safe enough for young children to be left alone, even for brief periods of time.
#4 It may be difficult for a short errand to be completed quickly enough to make leaving the children alone worthwhi
#5 It is **illegal** in some states to leave a child home alone for a certain period of time, or depending on their age.

Overview of Approach

leaving your children alone at home while running an errand

GPT
Reasons Prompt

REASONS

For		Against	
1.	....	1.	....
2.	....	2.	....
3.	....	3.	....
4.	....	4.	....
5.	....	5.	....

SBERT
Ave. Embedding

8	9	...	1	0

1024-D

Reasons for Leaving Young Children Alone at Home While Running an Errand:

#1 It is a good opportunity to give children a sense of **independence and responsibility**.
#2 Children can use it as an opportunity to study, exercise or practice a skill without interruption.
#3 It may help them **gain confidence** in their own abilities and trust in their parents.
#4 Errands can be done more efficiently without the need to supervise children's activities.
#5 It may be a **cheaper, more convenient** option than hiring a babysitter.

Reasons Against Leaving Young Children Alone at Home While Running an Errand:

#1 It is more difficult to monitor young children who are home alone and **ensure their safety**.
#2 It can be **emotionally and psychologically difficult** for some children who are not used to being alone.
#3 Some neighborhoods may not be safe enough for young children to be left alone, even for brief periods of time.
#4 It may be difficult for a short errand to be completed quickly enough to make leaving the children alone worthwhi
#5 It is **illegal** in some states to leave a child home alone for a certain period of time, or depending on their age.

Overview of Approach

leaving your children alone at home while running an errand

GPT
Reasons Prompt

REASONS

For		Against	
1.	....	1.	....
2.	....	2.	....
3.	....	3.	....
4.	....	4.	....
5.	....	5.	....

SBERT
Ave. Embedding

8	9	...	1	0

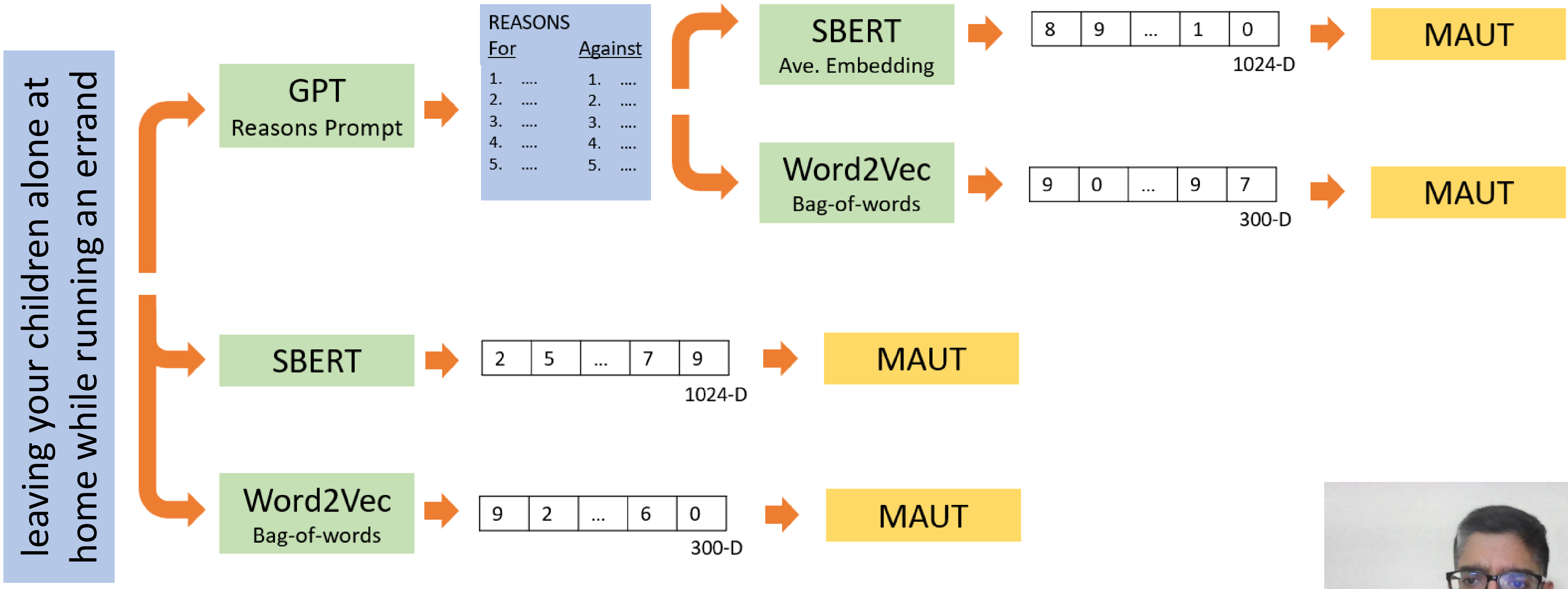
1024-D

MAUT

<u>Reasons for Leaving Young Children Alone at Home While Running an Errand:</u>
#1 It is a good opportunity to give children a sense of **independence and responsibility**.
#2 Children can use it as an opportunity to study, exercise or practice a skill without interruption.
#3 It may help them **gain confidence** in their own abilities and trust in their parents.
#4 Errands can be done more efficiently without the need to supervise children's activities.
#5 It may be a **cheaper, more convenient** option than hiring a babysitter.

<u>Reasons Against Leaving Young Children Alone at Home While Running an Errand:</u>
#1 It is more difficult to monitor young children who are home alone and **ensure their safety**.
#2 It can be **emotionally and psychologically difficult** for some children who are not used to being alone.
#3 Some neighborhoods may not be safe enough for young children to be left alone, even for brief periods of time.
#4 It may be difficult for a short errand to be completed quickly enough to make leaving the children alone worthwhi
#5 It is **illegal** in some states to leave a child home alone for a certain period of time, or depending on their age.

Overview of Approach



Study 1

150 participants recruited from Prolific Academic

Study 1

150 participants recruited from Prolific Academic

Extended DOSPERT scale with 150 items as well as additional demographic and psychographic items

Study 1

150 participants recruited from Prolific Academic

Extended DOSPERT scale with 150 items as well as additional demographic and psychographic items

MAUT fit on individual-level data with ridge regression

Study 1

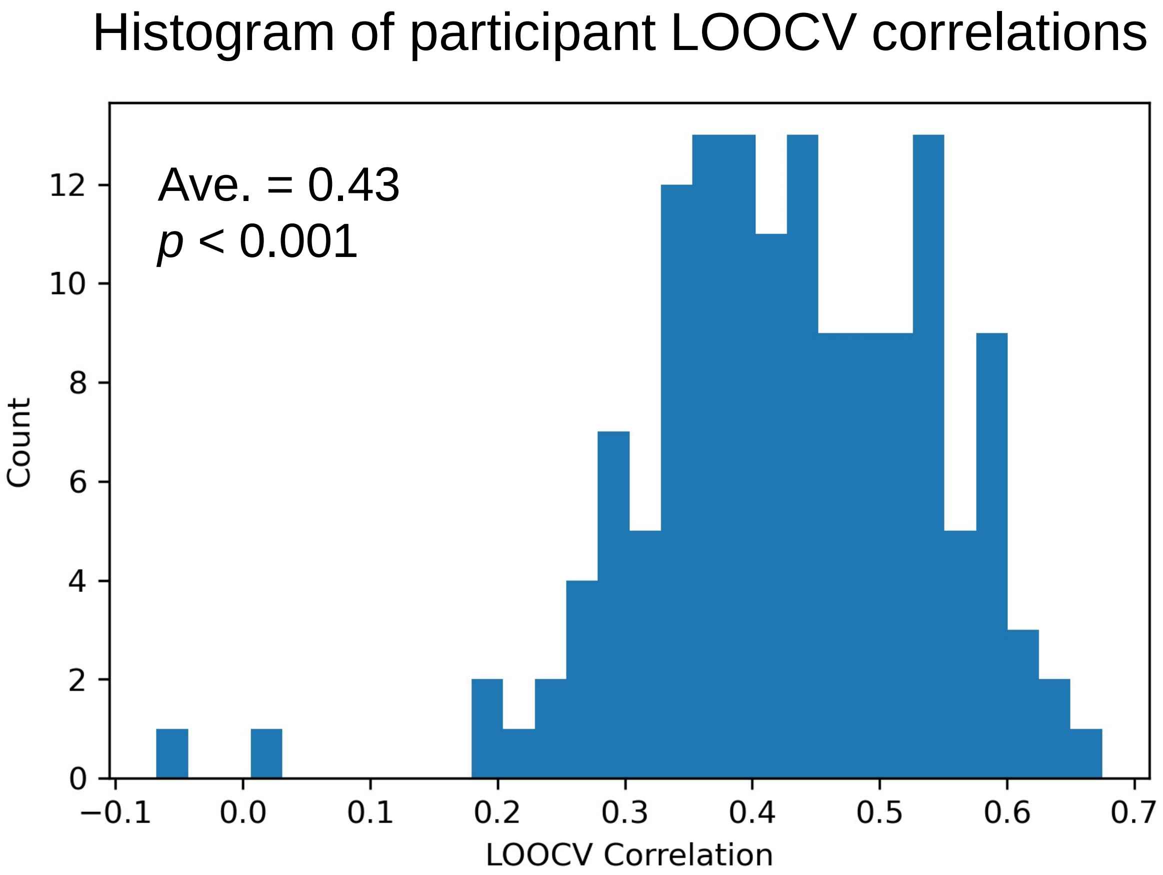
150 participants recruited from Prolific Academic

Extended DOSPERT scale with 150 items as well as additional demographic and psychographic items

MAUT fit on individual-level data with ridge regression

Preregistered at https://osf.io/amves/ and code and data available at https://osf.io/3y6ku/

Predictive Accuracy



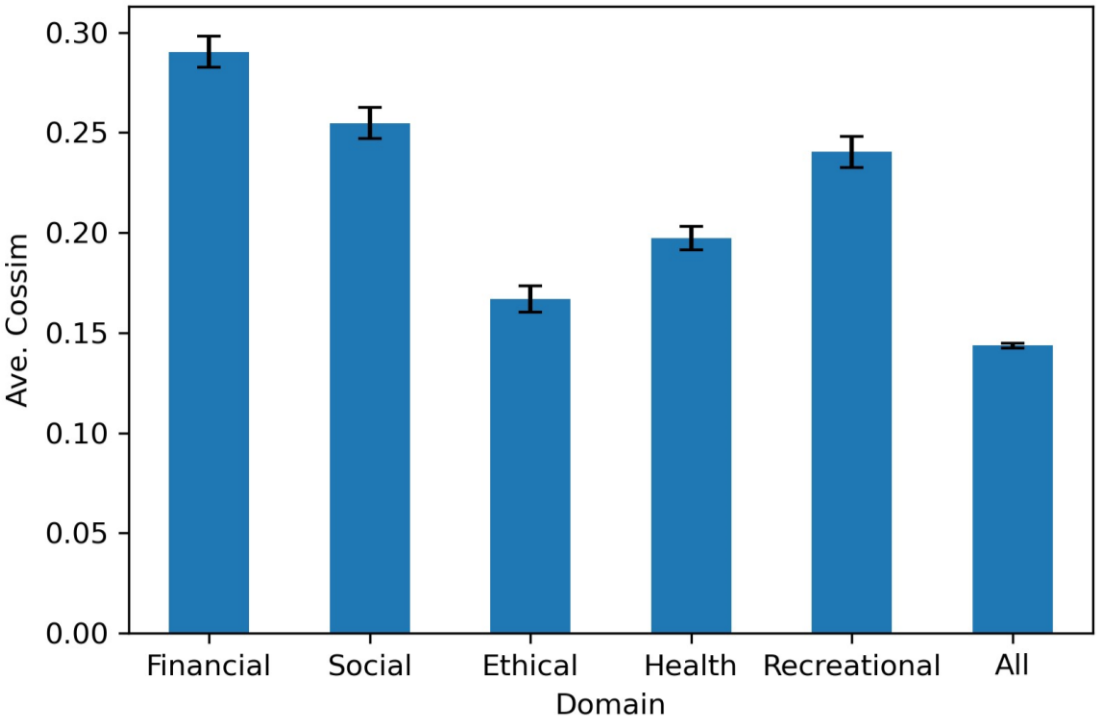
<u>Conclusions:</u>

Models make good predictions for most participants

This is also true when you use original DOSPERT items as training data and new items as test data (not shown here)

Item Variability

Similarity of items within DOSPERT domains

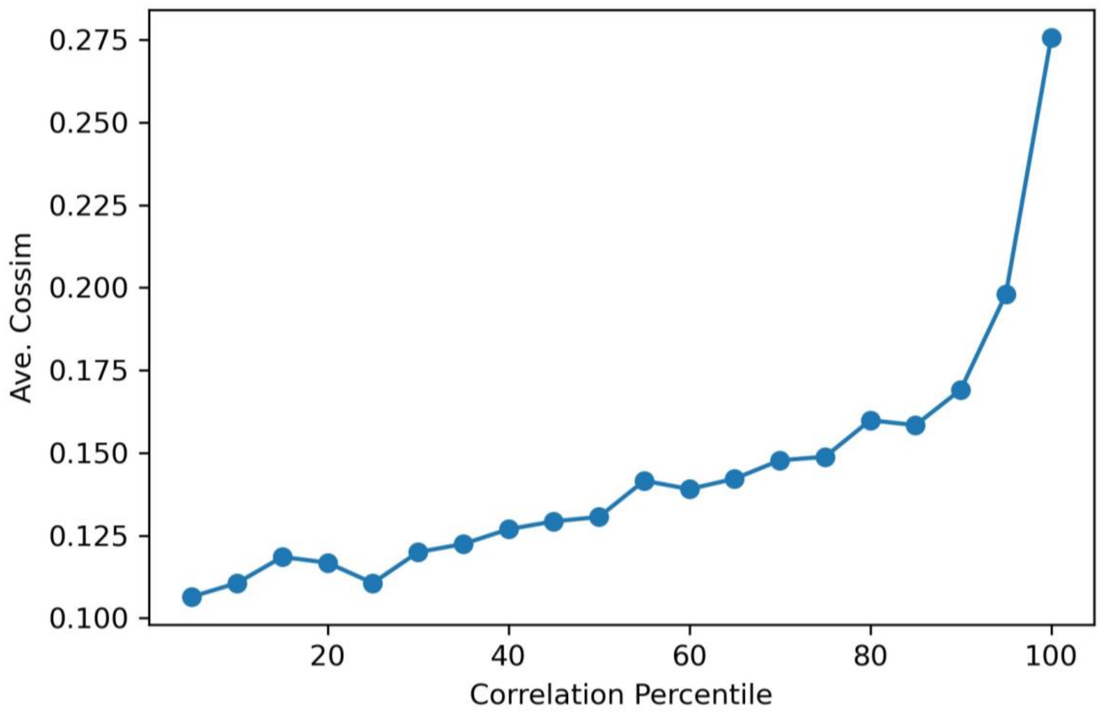


<u>Conclusions:</u>

LLM representations for DOSPERT items are more similar to items within domain than across domain

Item Variability

LLM item similarity vs. empirical correlation



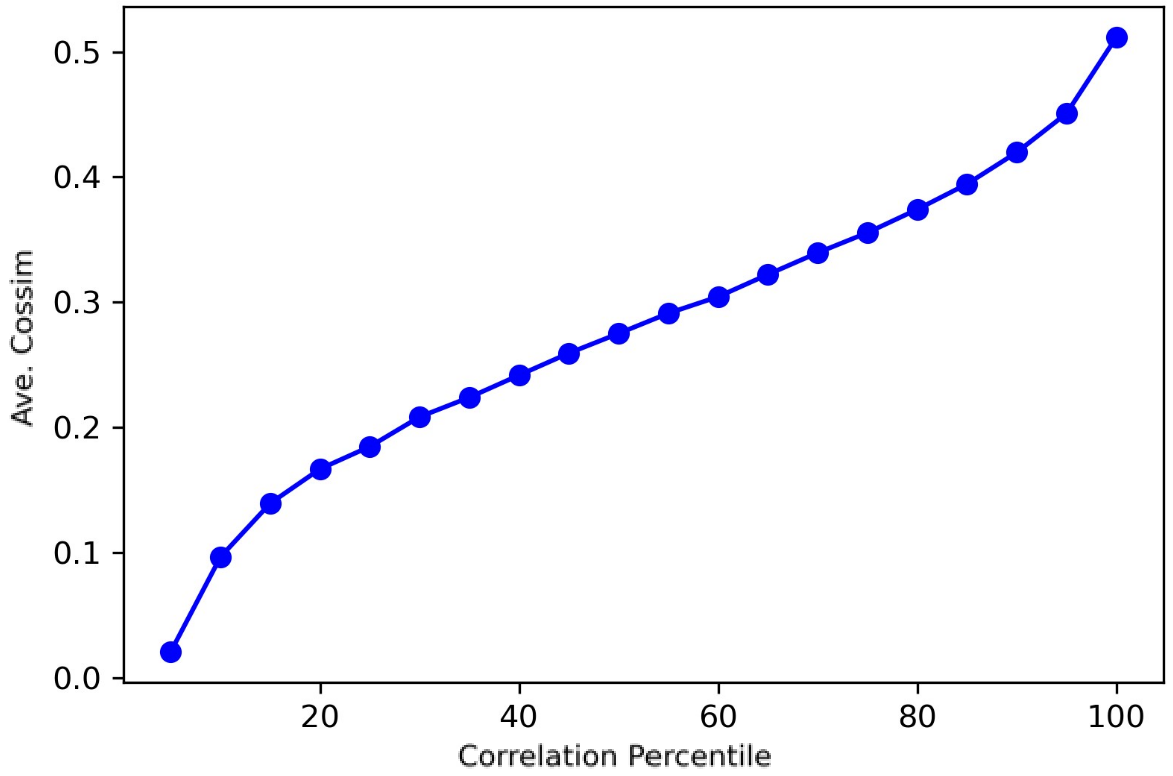
<u>Conclusions:</u>

Items that are highly correlated in data are also have similar LLM representations

Note that this relationship also holds within each DOSPERT domain (not shown here)

Participant Variability

MAUT weight similarity vs. empirical correlation



Conclusions:

Participants that give highly correlated ratings also have highly similar attribute weights

Study 2

150 participants recruited from Prolific Academic

Study 1 materials as well as **self-reported reasons** for five DOSPERT items

MAUT fit on individual-level data with ridge regression

Preregistered at https://osf.io/amves/ and code and data available at https://osf.io/3y6ku/

Study 2

150 participants recruited from Prolific Academic

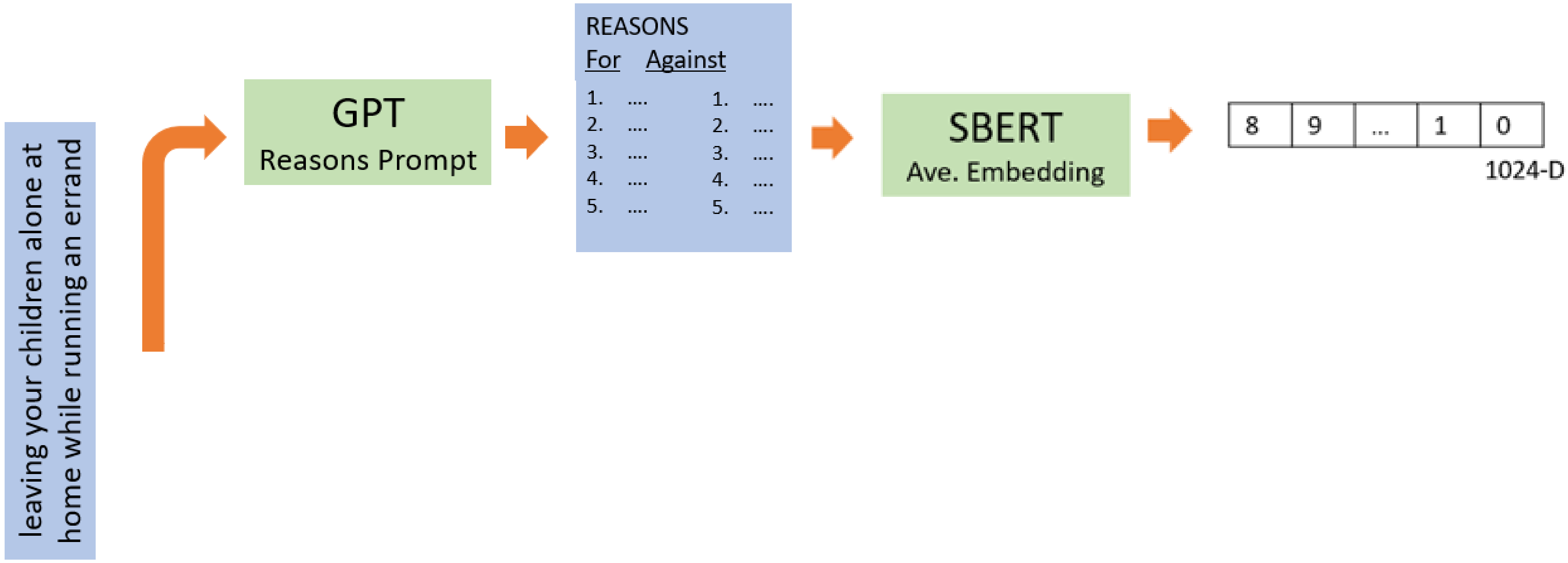
Study 1 materials as well as **self-reported reasons** for five DOSPERT items

MAUT fit on individual-level data with ridge regression

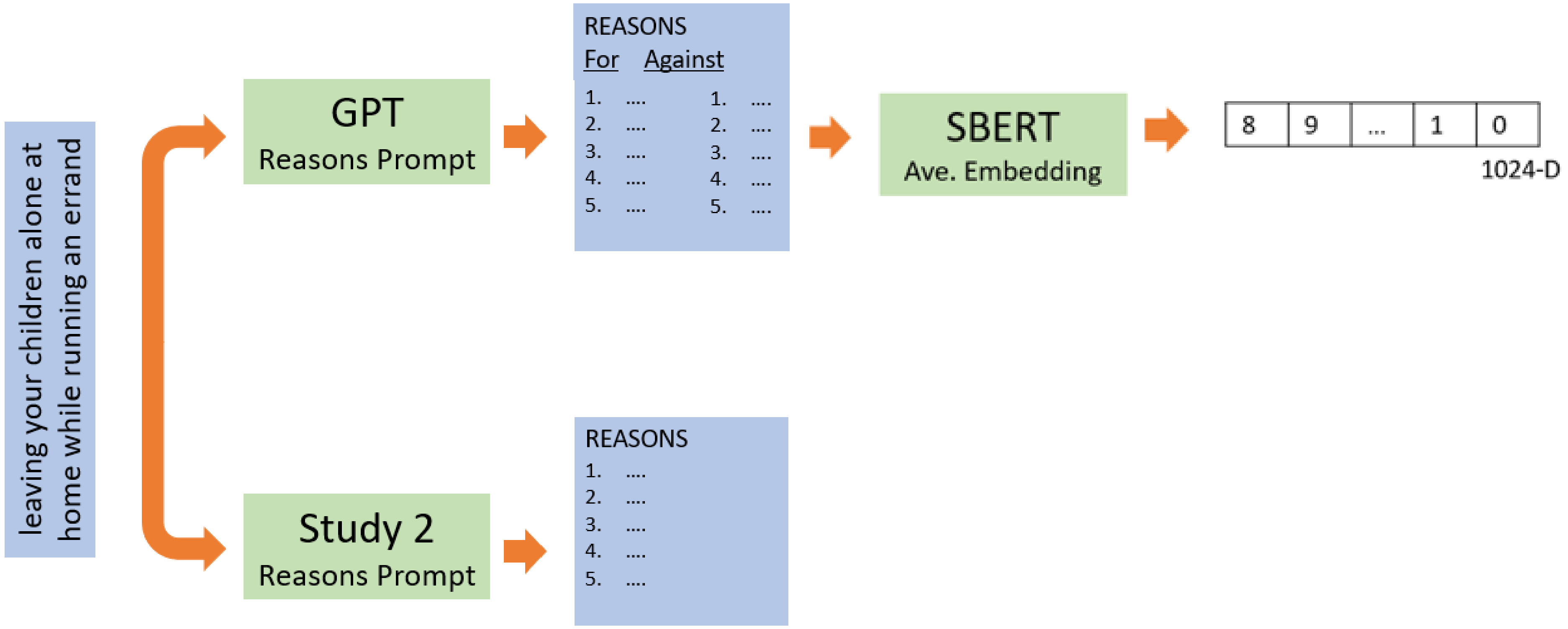
Preregistered at https://osf.io/amves/ and code and data available at https://osf.io/3y6ku/

All results from Study 1 replicate

Study 2

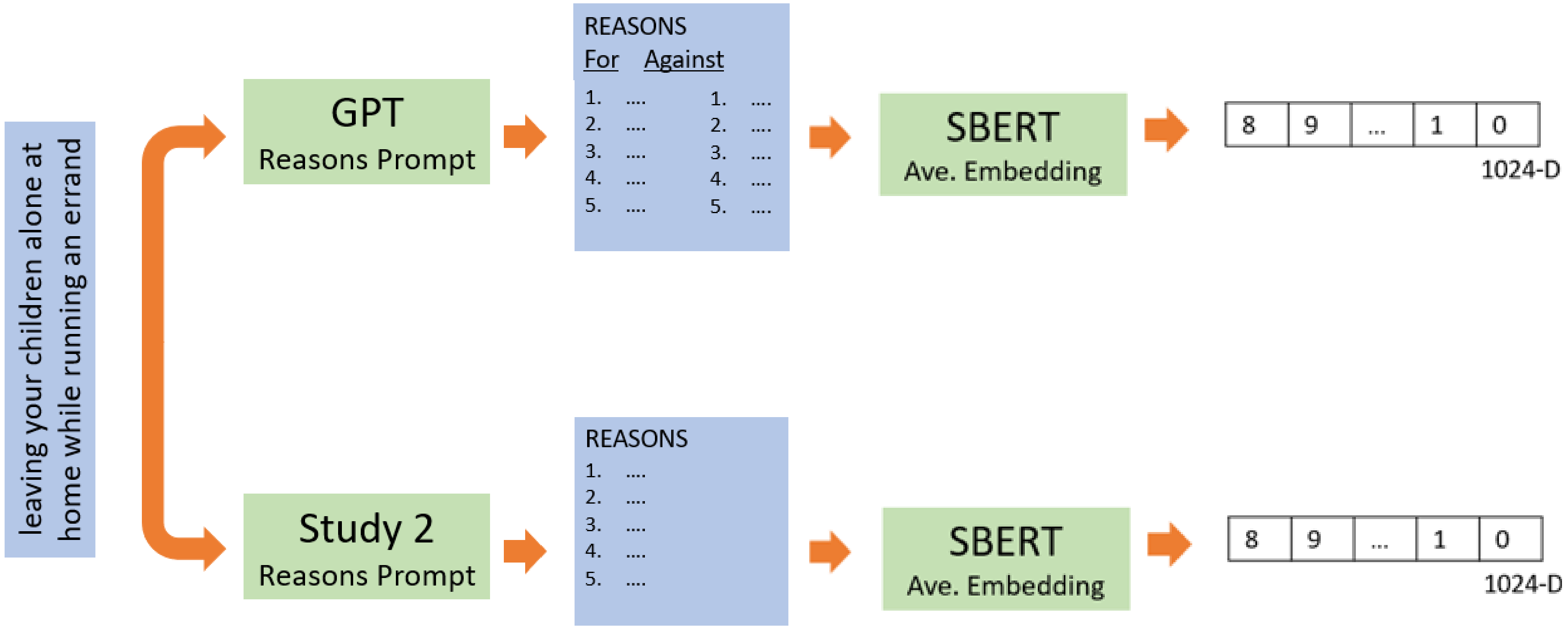

leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D

Study 2

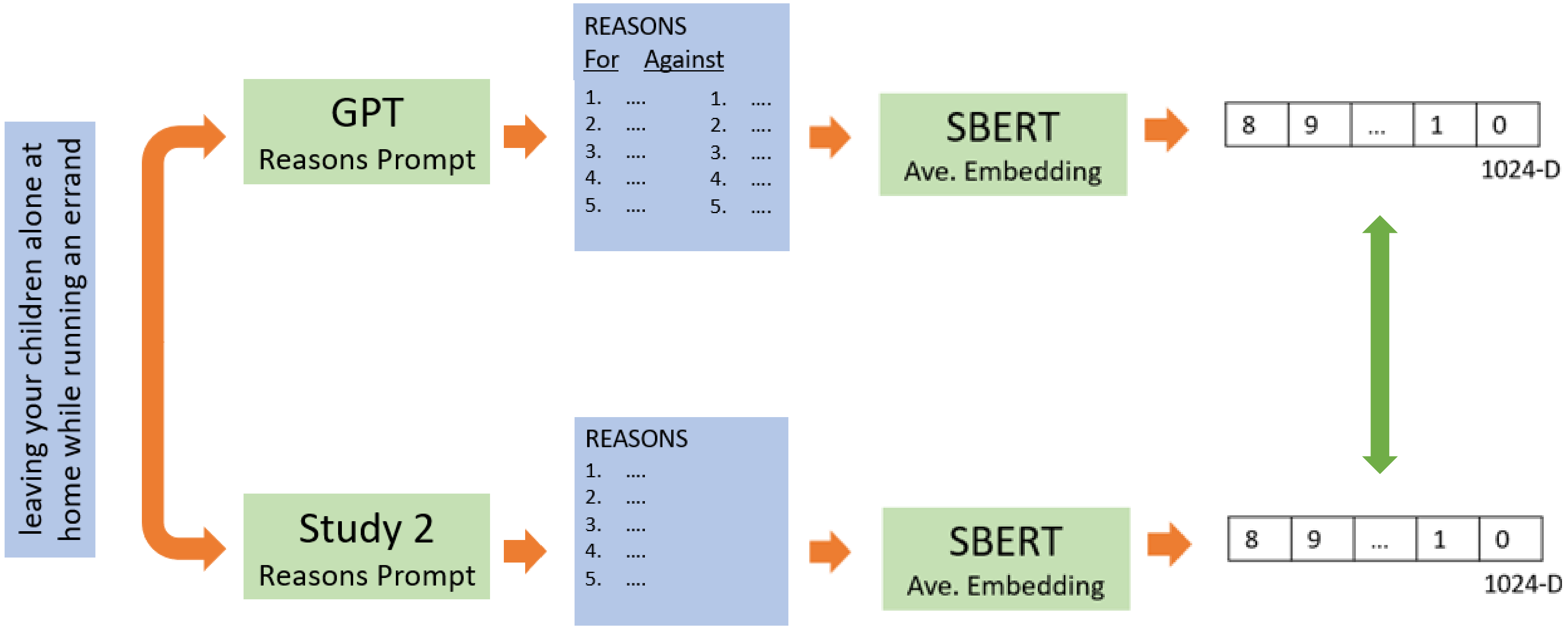


leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
Study 2
Reasons Prompt
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....

Study 2


leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
Study 2
Reasons Prompt
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D

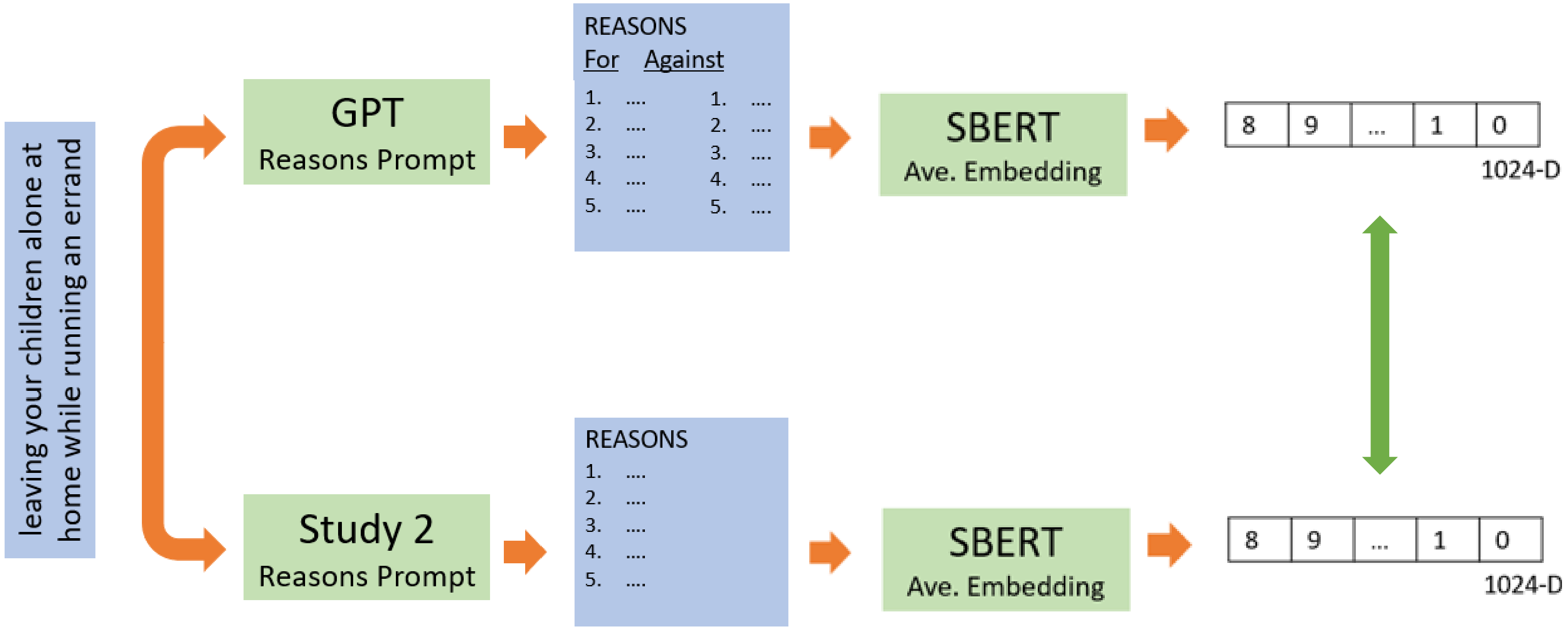
Study 2



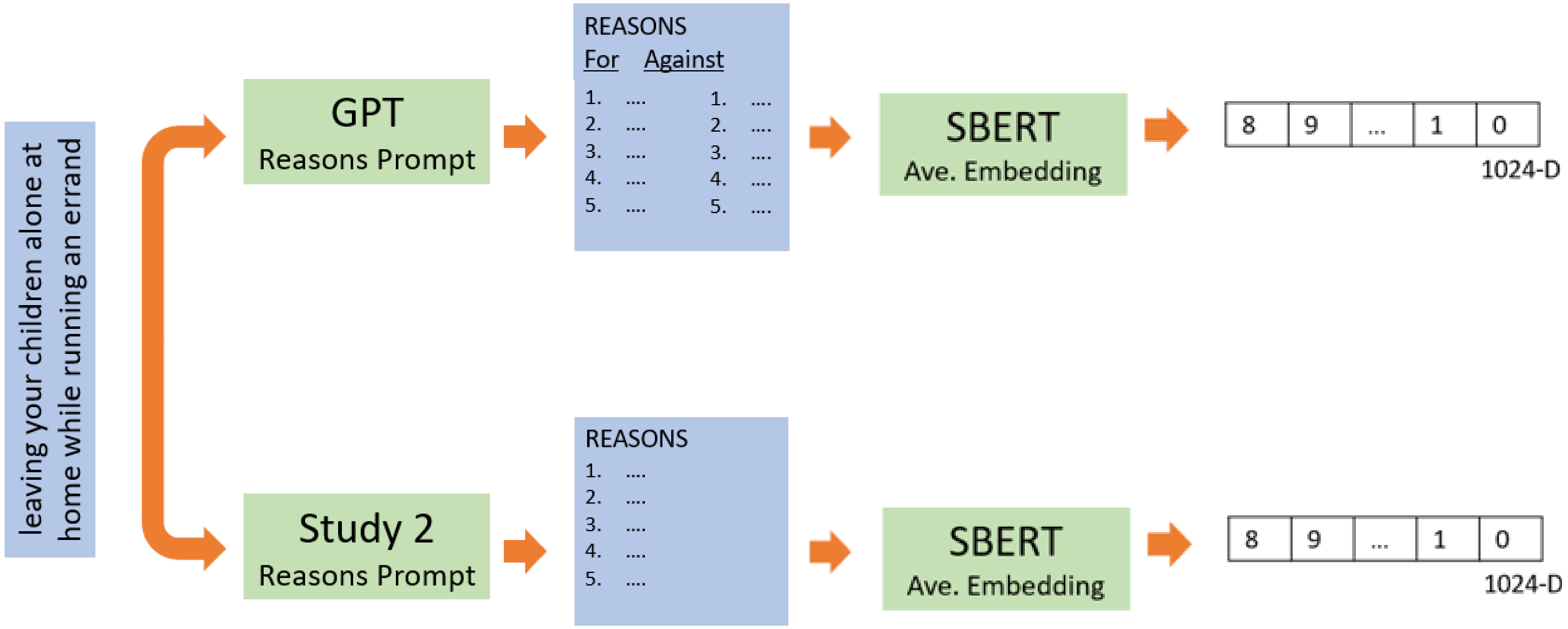
leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
Study 2
Reasons Prompt
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D

Study 2

More similar to each other than 98.1% of reasons for other items ($p < 0.001$)
(if random we would expect 50%)

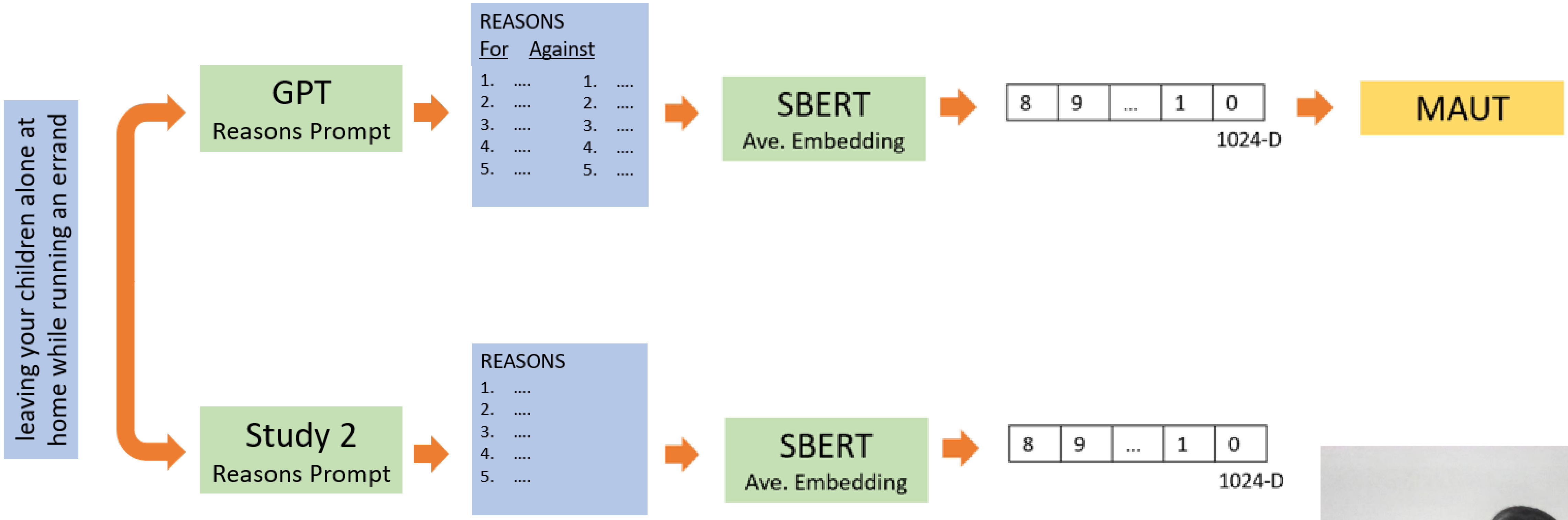


Study 2

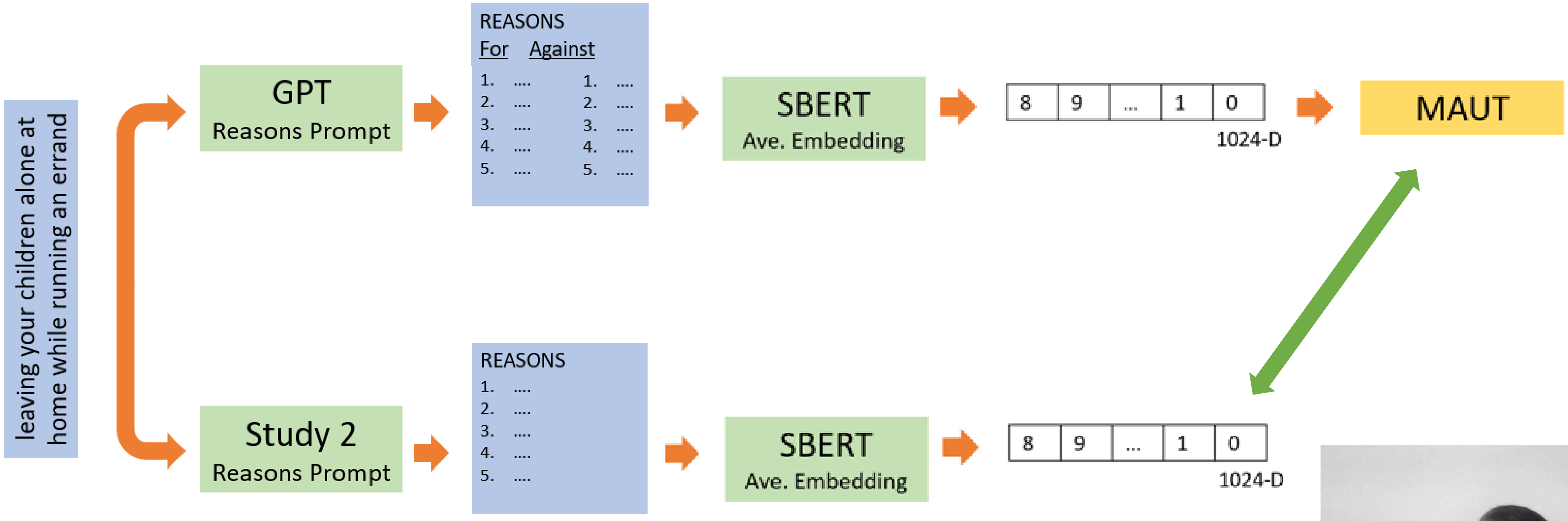


leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
Study 2
Reasons Prompt
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D

Study 2

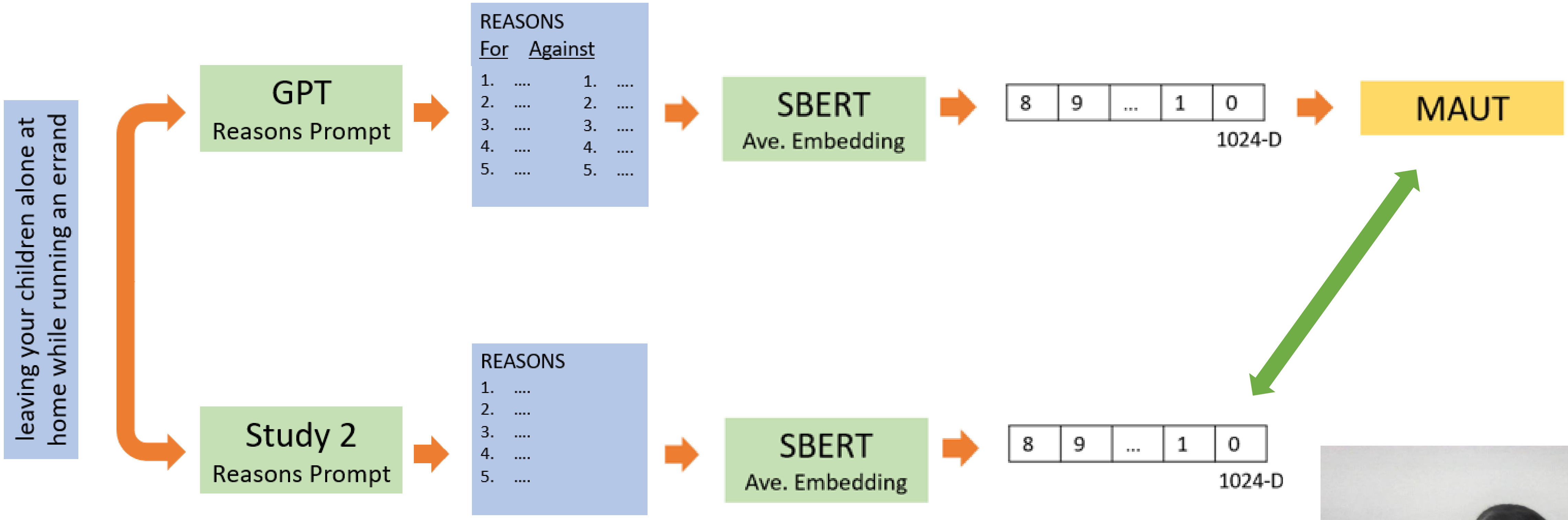

leaving your children alone at home while running an errand
GPT
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
MAUT
Study 2
Reasons Prompt
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D

Study 2


leaving your children alone at home while running an errand
GPT
Reasons Prompt
Study 2
Reasons Prompt
REASONS
For Against
1. .... 1. ....
2. .... 2. ....
3. .... 3. ....
4. .... 4. ....
5. .... 5. ....
REASONS
1. ....
2. ....
3. ....
4. ....
5. ....
SBERT
Ave. Embedding
SBERT
Ave. Embedding
8 9 ... 1 0
1024-D
8 9 ... 1 0
1024-D
MAUT

Study 2

More similar to each other than 64.3% of reasons of other participants (p < 0.001)
(if random we would expect 50%)



Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages
- Naturalistic items (e.g. common risks, behaviors, attitudes)

Disadvantages
- Agnostic about underlying decision process
- Cannot make out-of-sample predictions

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:
- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Disadvantages
- Artificial stimuli

How to synthesize?

Answer: Large language models!

Approach:

- LLMs quantify attributes of survey items
- MAUT weighs and aggregate LLM attributes
- Application to everyday risk taking

Study 1:

- Predict out-of-sample responses
- Predict variability in items
- Predict variability in individuals

Study 2:

- Predict subject-generated reasons

Psychometrics

Stimuli is lexical/linguistic (e.g. DOSPERT, Big-5)

Correlations in data reveal structure of variance

Advantages
- Naturalistic items (e.g. common risks, behaviors, attitudes)

Disadvantages
- Agnostic about underlying decision process
- Cannot make out-of-sample predictions

Decision Modeling

Stimuli is quantitative (e.g. attribute structures, gambles)

Decision models reveal structure of variance

Advantages:
- Modeling of decision processes (heuristics, context etc.)
- Out-of-sample predictions

Disadvantages
- Artificial stimuli

Synthesis:

- LLMs describe linguistic items as multi-attribute vectors
- Decision models weigh and aggregate LLM attributes
- Application to DOSPERT

Study 1:

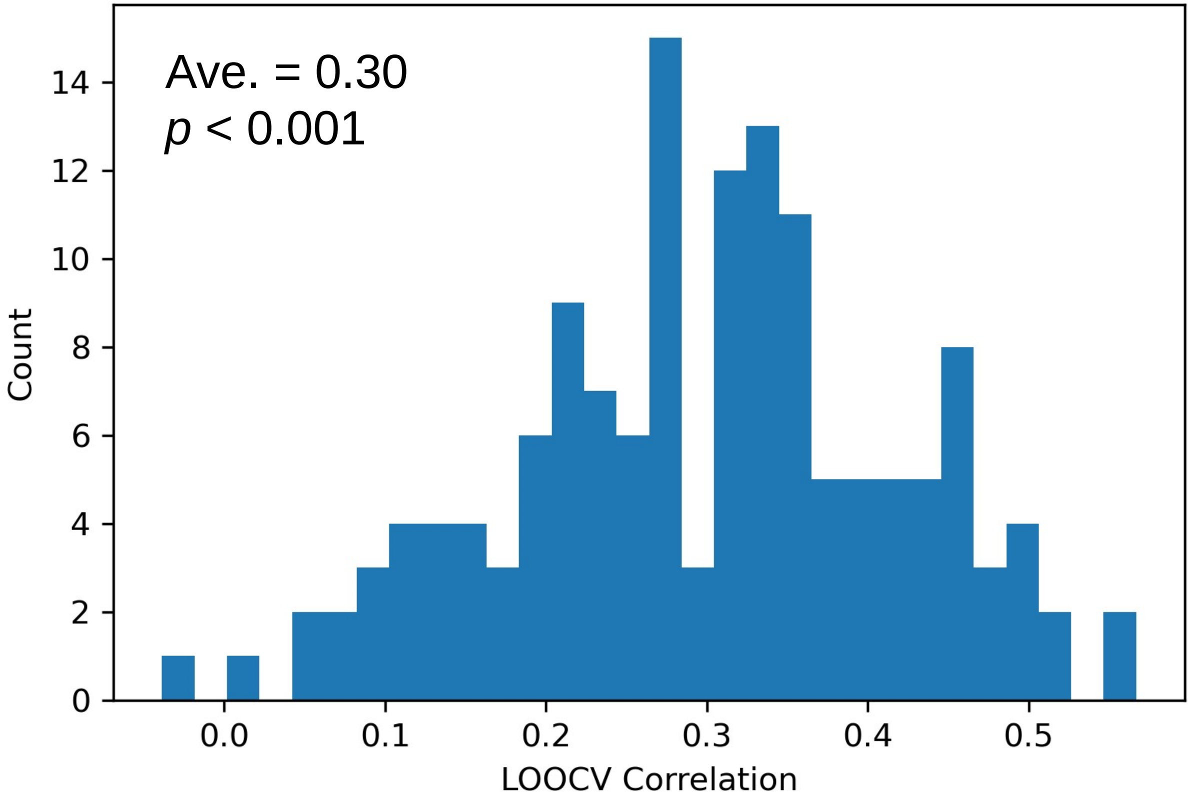
- Predict out-of-sample responses
- Predict correlations between items and individuals

Study 2:

- Predict subject-generated reasons

Predictive Accuracy

Participant-level correlations when fit on original DOSPERT items and tested on new items

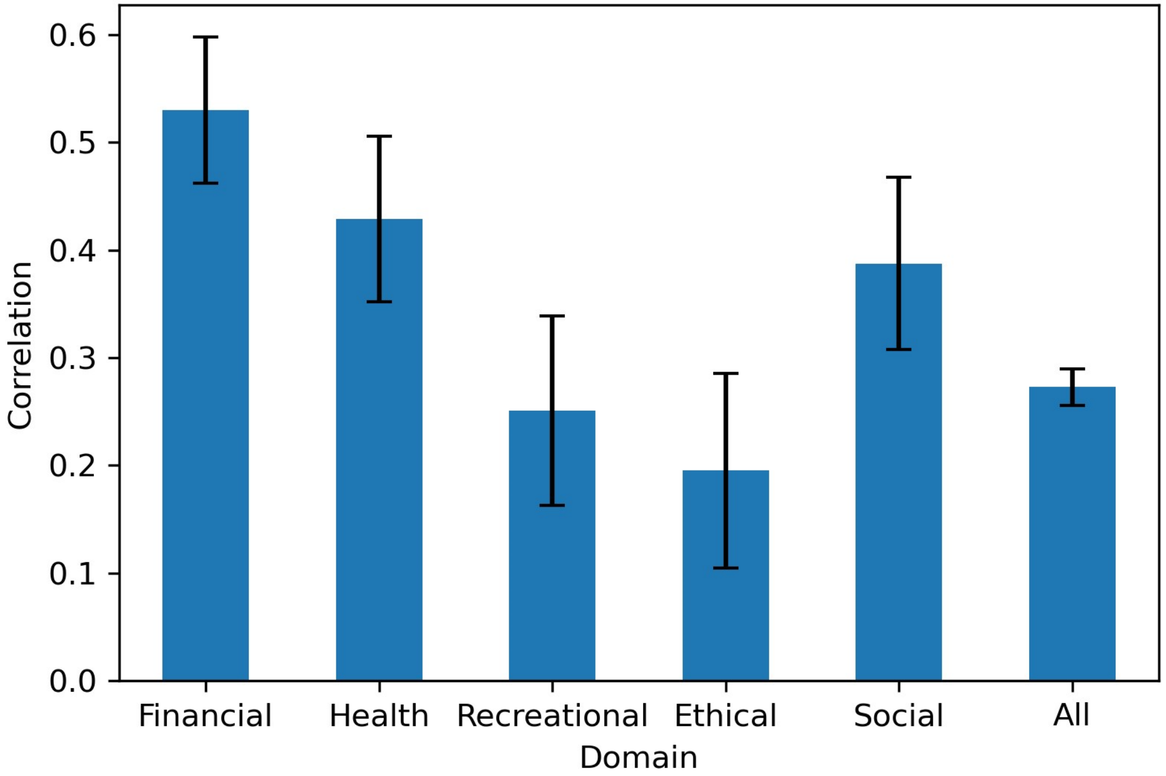


Results:

Models can be trained on smaller datasets, though accuracy drops somewhat

Item Variability

LLM similarity vs. empirical correlation for items



Related Work

Word representations can be used to model:

1. System-1 judgment (Bhatia, 2017, Psychological Review)
2. Multi-attribute choice (Bhatia & Stewart, 2018, Cognition)
3. Risk perception (Bhatia, 2019, Management Science)
4. Political judgment (Bhatia et al., 2019, SPPS)
5. Leadership perception (Bhatia et al., 2021, Leadership Quarterly)
6. Numerical judgment (Zou & Bhatia, 2021, Cognition)
7. Historical gender bias (Bhatia & Bhatia, 2021, Psychology of Women Quarterly)
8. Similarity judgment (Richie & Bhatia, 2022, Cognitive Science)
9. Food judgment (Gandhi et al., 2022, Psychological Science)
10. Memorability of words (Aka et al., 2023, Cognition)
11. Implicit attitudes (Bhatia & Walasek, 2023, PNAS)

Related Work

Sentence representations can be used to model:

11. Disease perception (Aka & Bhatia, 2021, JACR)
12. Behavioral propensity (Singh et al., 2021, CB&B)
13. Reason-based choice (Zhao et al., 2022, Psychological Review)
14. Semantic verification (Bhatia & Richie, in press, Psychological Review)
15. Personality prediction (Abdurahman et al., in press, JPSP)

Related Work

Sentence representations can be used to model:

11. Disease perception (Aka & Bhatia, 2021, JACR)
12. Behavioral propensity (Singh et al., 2021, CB&B)
13. Reason-based choice (Zhao et al., 2022, Psychological Review)
14. Semantic verification (Bhatia & Richie, in press, Psychological Review)
15. Personality prediction (Abdurahman et al., in press, JPSP)

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Current Opinion in Behavioral Sciences

Distributed semantic representations for modeling human judgment

Sudeep Bhatia, Russell Richie and Wanling Zou

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE
Current Directions in Psychological Science
1–8

DOI: 10.1177/09637214211068113
www.psychologicalscience.org/CDPS
SAGE

Cognitive Modeling With Representations From Large-Scale Digital Data

Sudeep Bhatia and Ada Aka
Department of Psychology and Department of Marketing, University of Pennsylvania

Related Work

Sentence representations can be used to model:

11. Disease perception (Aka & Bhatia, 2021, JACR)
12. Behavioral propensity (Singh et al., 2021, CB&B)
13. **Reason-based choice** (Zhao et al., 2022, Psychological Review)
14. Semantic verification (Bhatia & Richie, in press, Psychological Review)
15. **Personality prediction** (Abdurahman et al., in press, JPSP)

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Current Opinion in Behavioral Sciences

Distributed semantic representations for modeling human judgment

Sudeep Bhatia, Russell Richie and Wanling Zou

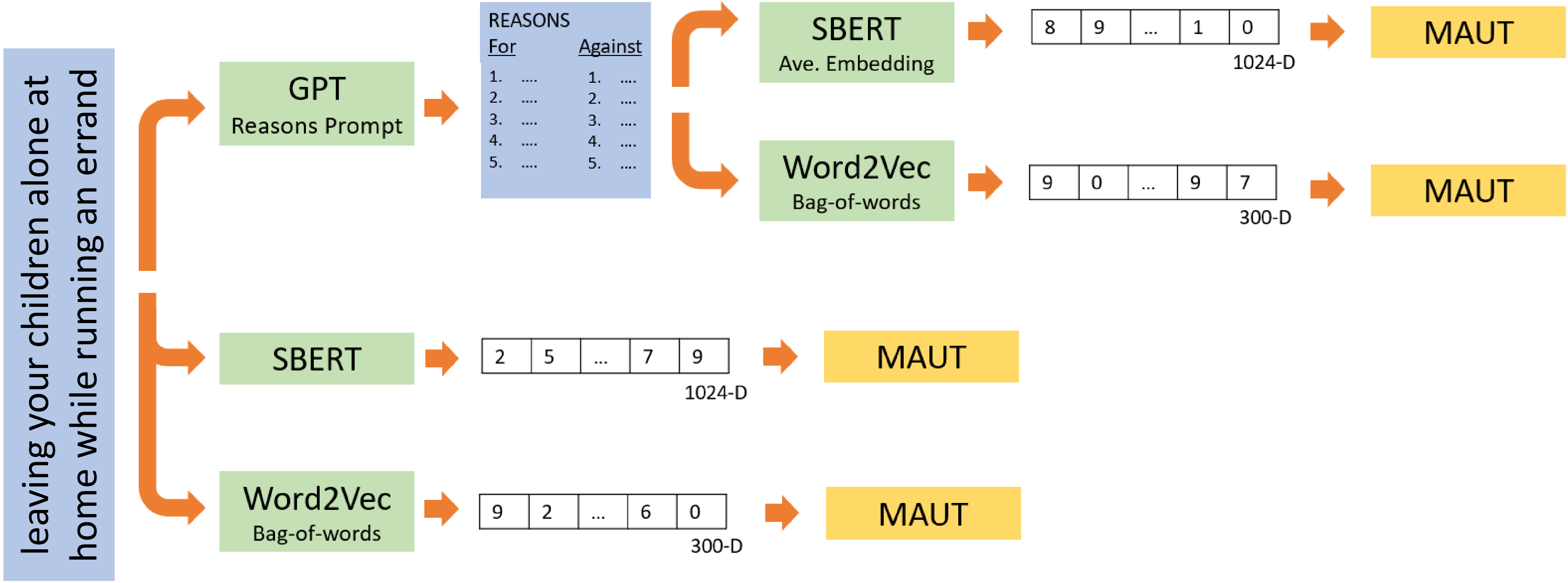
aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE
Current Directions in Psychological Science
1–8

DOI: 10.1177/09637214211068113
www.psychologicalscience.org/CDPS
SAGE

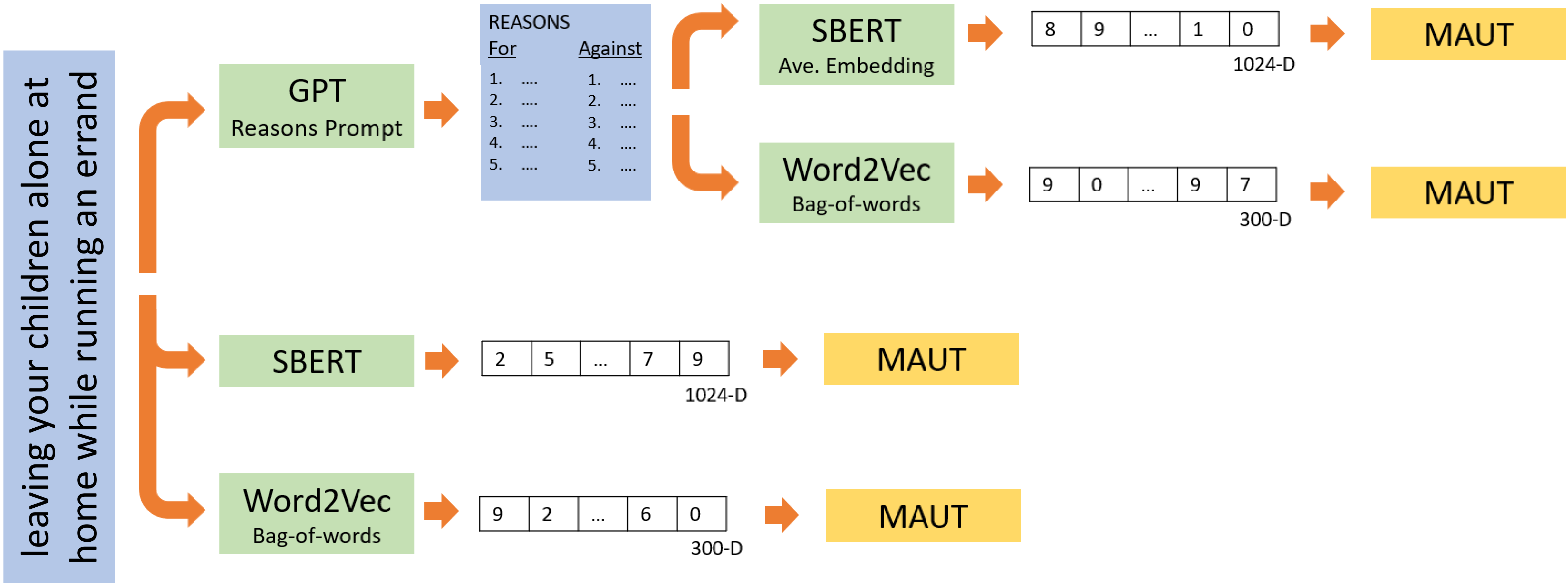
Cognitive Modeling With Representations From Large-Scale Digital Data

Sudeep Bhatia and Ada Aka
Department of Psychology and Department of Marketing, University of Pennsylvania

Overview of Approach

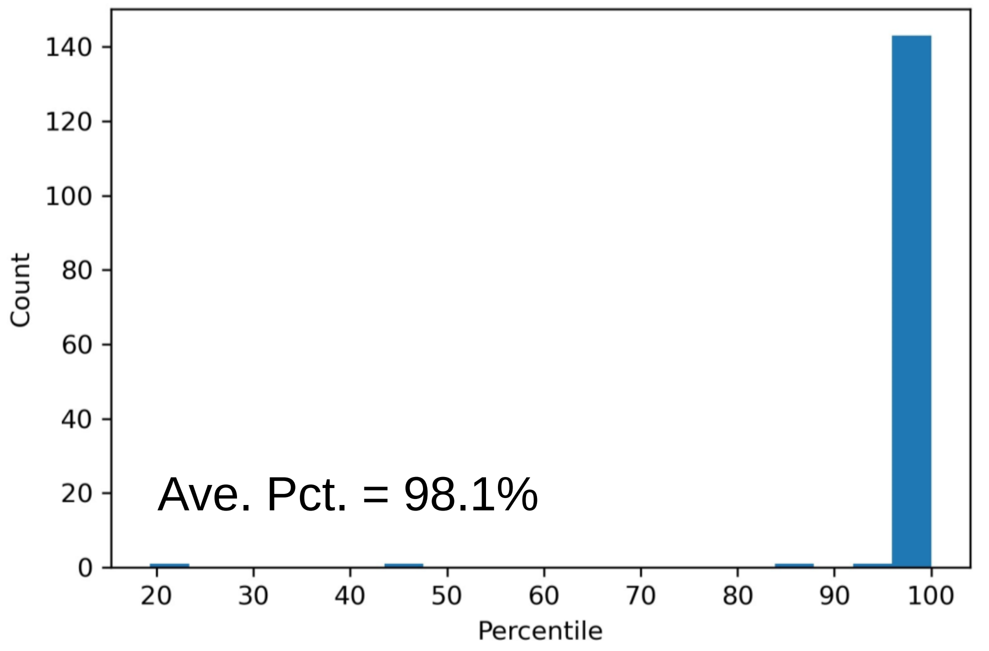


Overview of Approach



Item Variability

Percentile ranks for similarity of GPT-generated reasons to human-generated reasons for an item, relative to human-generated reasons for other items

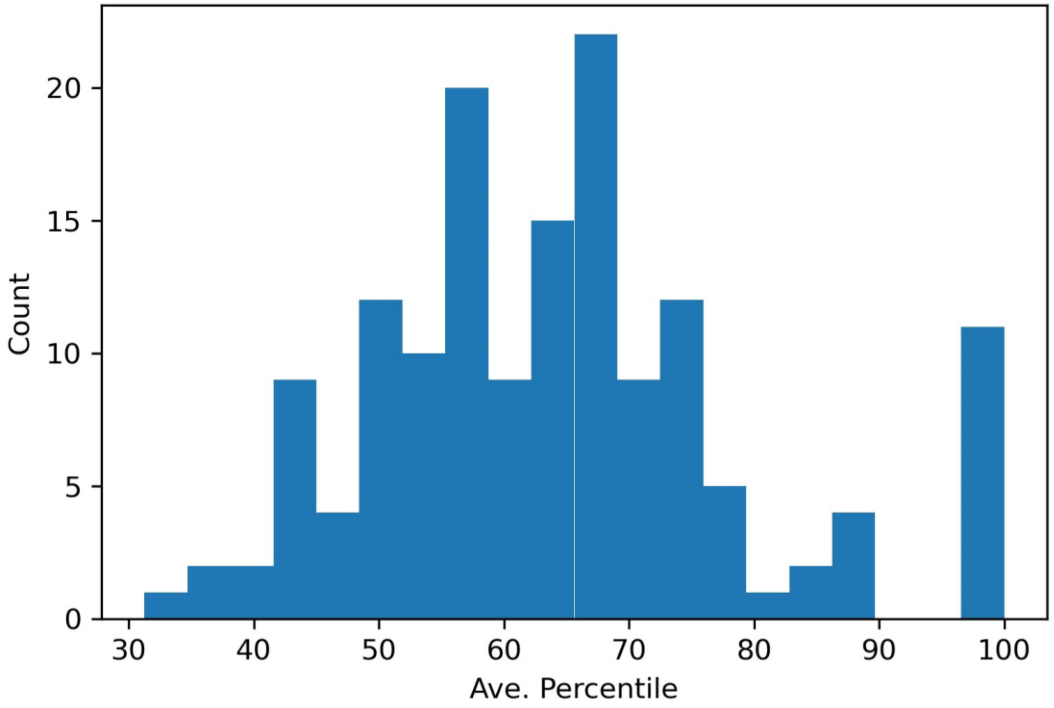


Conclusions:

GPT-generated reasons for an item are nearly always more similar to human-generated reasons for an item relative to human-generated reasons for other items

Individual Variability

Percentile ranks for similarity of GPT-generated reasons are to a specific individual's reasons for an item, relative to other individuals' reasons for that item



<u>Conclusions:</u>

Attribute dimensions identified by best-fitting MAUT model to a participant are reflected in that participant's reasons for that item