Neural representations of Abstract Concepts across English and Mandarin: Similar Neural Infrastructure with Differing Utilization

Robert Vargas and Marcel Adam Just

Carnegie Mellon University, Center for Cognitive Brain Imaging

Introduction

Motivation: To determine the representational commonality and differences in the neural representations of abstract concepts (such as *truth*) between native English and Mandarin speakers using MVPA applied to fMRI.

Issue: The aperceptual nature of abstract concepts brings into question the commonality of their neural representations across cultures and languages**.** Recent work examining the neural representation of abstract concepts in monolingual native English speakers suggests that there exists at least 3 underlying semantic dimensions associated with the human processing of abstract concepts (Vargas and Just, 2019). This study measured the commonality of the neural representation of individual abstract concepts and the degree to which these hypothesized underlying semantic dimensions generalize to a sample of native Mandarin speakers.

Primary findings:

- Factor analysis of the activation patterns in both languages reveal 4 language-invariant underlying semantic dimensions associated with the neural representation of abstract concepts. These semantic dimensions suggest a culturally invariant network of regions underlying the structure of the thinking of abstract concepts.
- Classification analyses revealed concept-level commonalities and differences across languages.

Methods

English sample: 10 adults (6 females; 20-38 years old; M = 25.89)
- 1 subject was excluded from analyses due to falling asleep.

Mandarin sample: 10 adults (5 females; 18-26 years old; M = 21.2),
- 3 subjects were excluded from analysis due to scanning related issues
- Native Mandarin speakers who have spent less than 1 year living outside PRC

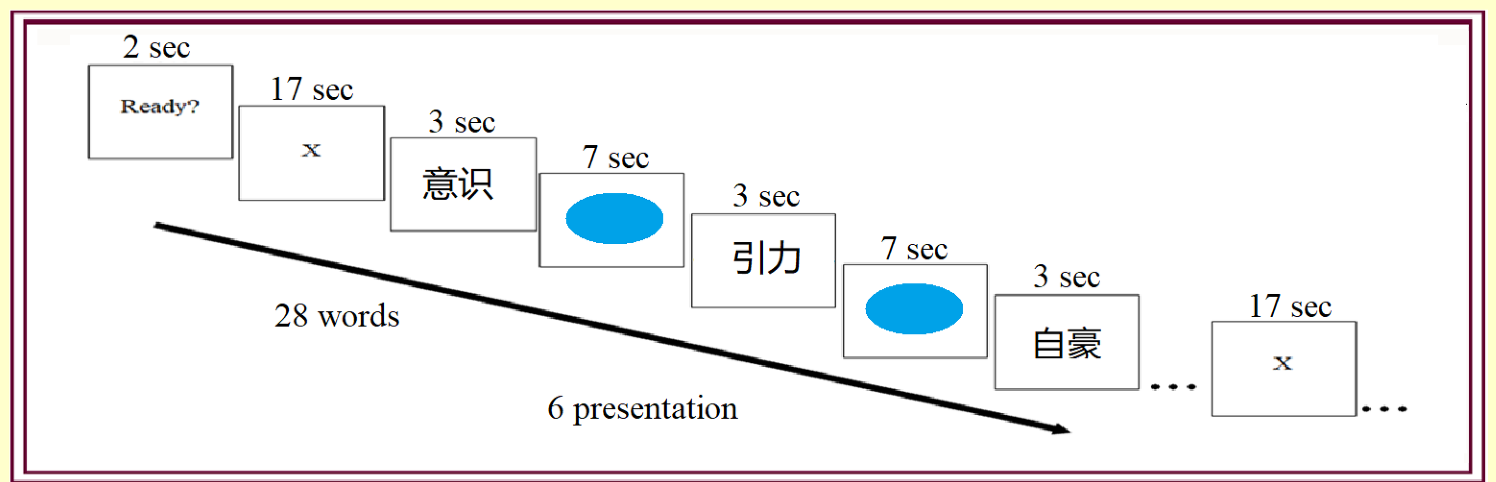
Stimuli

- 28 abstract concepts present in both studies. The 7 semantic category labels presented here for clarity.)

Math	Physics	Social	Emotion	Law	Metaphysical	Causality
subtraction (减法)	gravity (引力)	gossip (绯闻)	happiness (幸福)	contract (合同)	causality (因果关系)	deity (神明)
equality (相等)	force (力)	intimidation (恐吓)	sadness (悲伤)	ethics (道德)	consciousness (意识)	spirituality (灵性)
probability (概率)	heat (热能)	forgiveness (谅解)	anger (愤怒)	crime (罪行)	truth (真理)	sacrilege (亵渎)
multiplication (乘法)	acceleration (加速度)	compliment (赞美)	pride (自豪)	exoneration (免罪)	necessity (必要性)	faith (信仰)

Stimulus Presentation

- Participants were instructed to think about the main properties of each word concept when it was presented in the scanner.
- It was emphasized that they should think of the same properties each time a given word was presented.



Data Analysis Methods

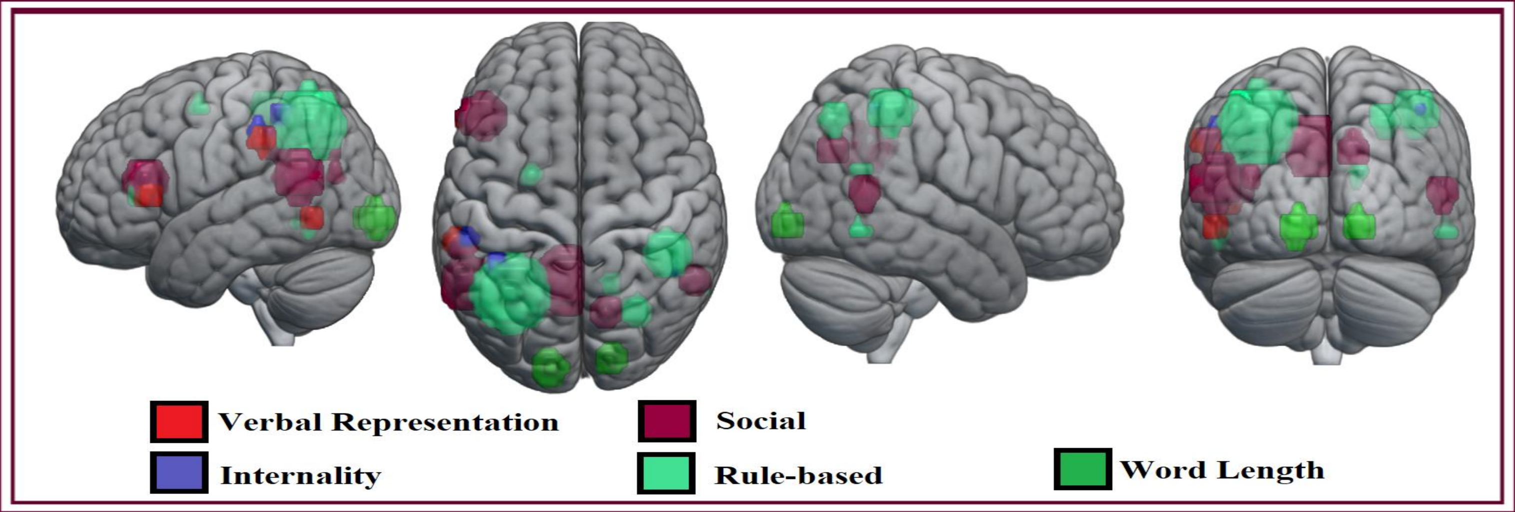
Classification Analyses

- **Within-participant concept decoding: A** Gaussian Naïve Bayes (GNB) classifier was trained on the activation patterns in 4 of the 6 presentations of the 28 concepts and tested on the mean of the 2 left-out presentations.
- **Between-participant within language decoding**: A GNB classifier was trained on all but left out participant within their language group and was tested on the left out participant.
- **Cross language decoding:** To provide a measure of the similarity across languages, a GNB classifier was trained on all participants from one language and tested on each participant from the other language.

Factor Analysis and Predictive Modeling

- A two-level factor analysis was computed on the English participants (procedure described in detail in Just et al. 2014) revealing 3 group-level interpretable factors.
- An independent group of participants was asked to rate each of the 28 concepts on a scale from 1-7 with respect to its salience to each of the dimensions, as they were interpreted here (e.g. degree to which a concept, such as *faith*, is verbally versus perceptually based). These ratings were used in a predictive model to predict the activation patterns of abstract concepts for which the model had no activation data (Mitchell et al., 2008). Using the factor locations and activation data for all but one participant, a multiple regression model, using the behavioral ratings was computed to predict activation patterns for , individual concepts. Mean accuracy across concepts and participants is 0.73, $p<0.01$.

Results



Underlying Neural Semantic Dimensions (shown above)

Verbal Representation: This dimension refers to the degree to which a concept is verbally represented as opposed to visuospatially represented. *Faith*, *ethics* and *necessity* lie at the verbal extreme while *gravity*, *force*, and *heat* lie at the visuospatial extreme. The extreme items have extreme factor scores on that factor.

Internality: The degree to which a concept is experienced as a state or event that is external versus internal to **oneself.** An event that is external requires the representation of the world outside oneself and relative non-involvement of one's own state. Items found at the external extreme include *causality*, *sacrilege*, and *crime.* Items at the internal extreme include *spirituality*, *sadness*, and *pride*

Social Content: The degree to which a concept involves social interaction or self-perception as viewed in a social context. *Forgiveness*, *gossip*, and *intimidation* typify the this.

Rule-Based: This dimension organizes information that defines or is defined by specific, precise relationships between other concepts. This includes mathematical symbols or equations as well as concepts that are involved in establishing or being based on a set of rules. The concepts *multiplication*, *probability*, and *ethics* typify this dimension.

Word Length: This dimension captures the neural encoding of the size of the word, measured as the number of characters or strokes (depending on the language).

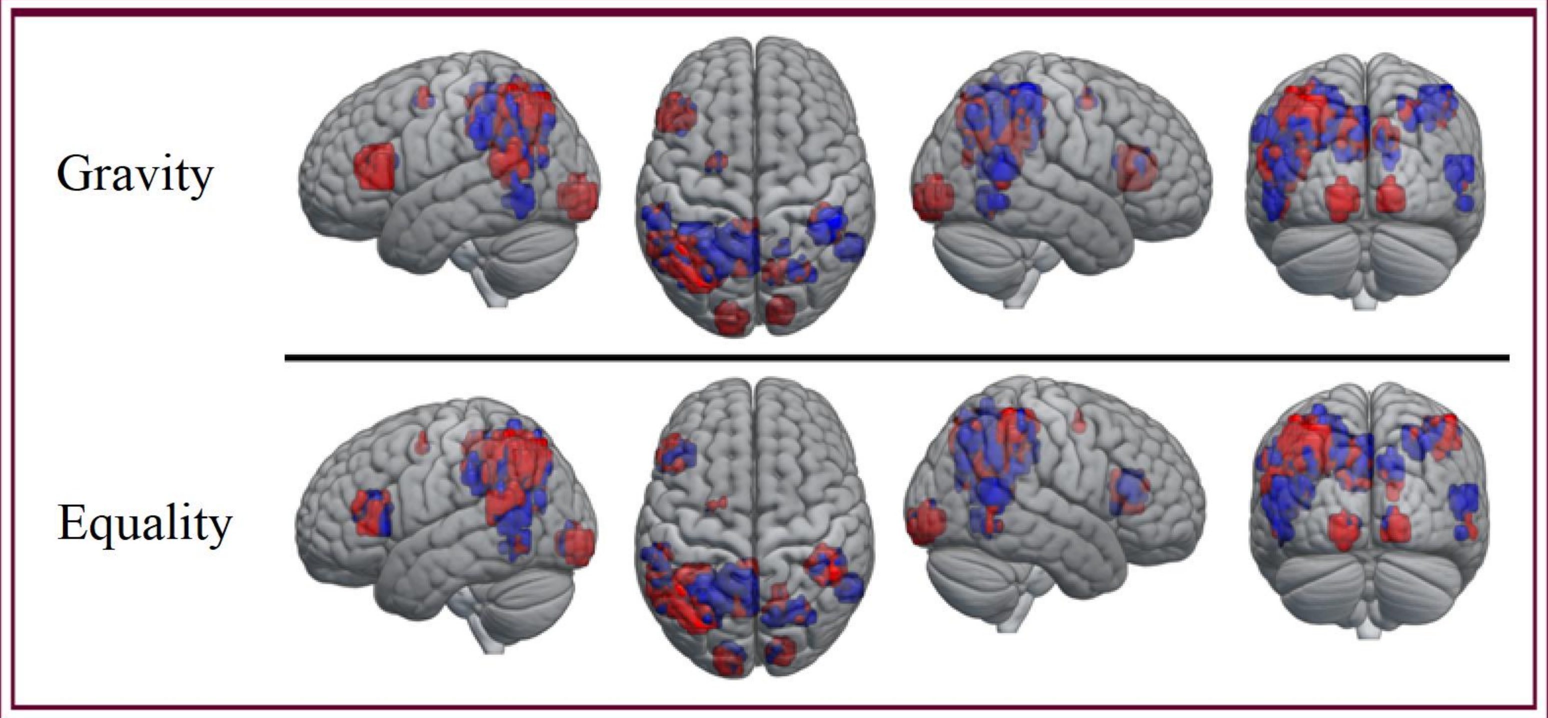
Accuracy of Decoding Abstract Concepts Within and Across languages

- Mean between-participant within language, shown in 3 left-most columns
- Mean within participant accuracies shown in parentheses
- Mean cross-language decoding (rightmost column)

	Within-Language Mandarin	Within-Language English	Within-Language Mean	Cross-Languge
subtraction	0.72 (0.79)	0.94 (0.89)	0.83	0.70
equality	0.7 (0.77)	0.58 (0.8)	0.64	0.67
probability	0.69 (0.72)	0.8 (0.86)	0.75	0.49
multiplication	0.91 (0.81)	0.73 (0.92)	0.82	0.82
gravity	0.8 (0.8)	0.86 (0.88)	0.83	0.79
force	0.84 (0.84)	0.84 (0.87)	0.84	0.83
heat	0.63 (0.7)	0.78 (0.8)	0.71	0.74
acceleration	0.92 (0.77)	0.76 (0.84)	0.84	0.77
gossip	0.63 (0.79)	0.66 (0.79)	0.65	0.68
intimidation	0.72 (0.82)	0.75 (0.83)	0.74	0.73
forgiveness	0.75 (0.69)	0.77 (0.7)	0.76	0.80
compliment	0.81 (0.72)	0.67 (0.81)	0.74	0.58
happiness	0.76 (0.71)	0.75 (0.76)	0.76	0.60
sadness	0.69 (0.78)	0.71 (0.85)	0.70	0.76
anger	0.59 (0.79)	0.62 (0.82)	0.61	0.62
pride	0.78 (0.86)	0.86 (0.88)	0.82	0.84
contract	0.52 (0.77)	0.63 (0.75)	0.58	0.66
ethics	0.71 (0.78)	0.72 (0.83)	0.72	0.56
crime	0.66 (0.72)	0.76 (0.8)	0.71	0.69
exoneration	0.63 (0.77)	0.76 (0.83)	0.70	0.41
causality	0.91 (0.89)	0.8 (0.89)	0.86	0.62
consciousness	0.66 (0.77)	0.79 (0.84)	0.73	0.53
truth	0.69 (0.77)	0.62 (0.86)	0.66	0.55
necessity	0.9 (0.82)	0.78 (0.77)	0.84	0.57
deity	0.77 (0.69)	0.59 (0.78)	0.68	0.60
spirituality	0.57 (0.72)	0.79 (0.82)	0.68	0.46
sacrilege	0.61 (0.73)	0.62 (0.83)	0.62	0.52
faith	0.81 (0.72)	0.78 (0.81)	0.80	0.61
Mean	**0.73**	**0.74**	**0.73**	**0.65**

Comparing individual concepts across languages

- Univariate GLM contrasts of each of the 3 most decodable (*gravity*, *force*, and *ethics*) and least decodable (*equality*, *intimidation*, and *causality*) concepts suggests that no single region accounted for the differences between English and Mandarin. For example, for English the concept equality a had higher activation in regions associated with social processing while in Mandarin, this concept showed higher activation levels in regions associated with verbal processing (shown below).



Conclusions

- A common neural infrastructure underlies the representation of abstract concepts in both languages.
- Concept-level differences in subregions within the shared factor locations suggest that the same neural infrastructure is relied on regardless of language but some regions are sometimes activated to different degrees in the two languages

References

- Vargas R, Just MA. 2019. Neural representations of Abstract Concepts: Identifying Underlying Neurosemantic Dimensions. *Cerebral Cortex*, 30(*4*), 2157-2166.
- Just MA, Cherkassky VL, Buchweitz A, Keller TA, Mitchell TM. 2014. Identifying autism from neural representations of social interactions: Neurocognitive markers of autism. PLoS ONE, 12, e113879
- Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. 2008. Predicting human brain activity associated with the meanings of nouns. Science. 5880:1191–1195.