

Disentangling Topic Models: A Cross-Cultural Analysis of Personal Values through Words



Steven R. Wilson and Rada Mihalcea
University of Michigan

Ryan L. Boyd and James W. Pennebaker
The University of Texas at Austin

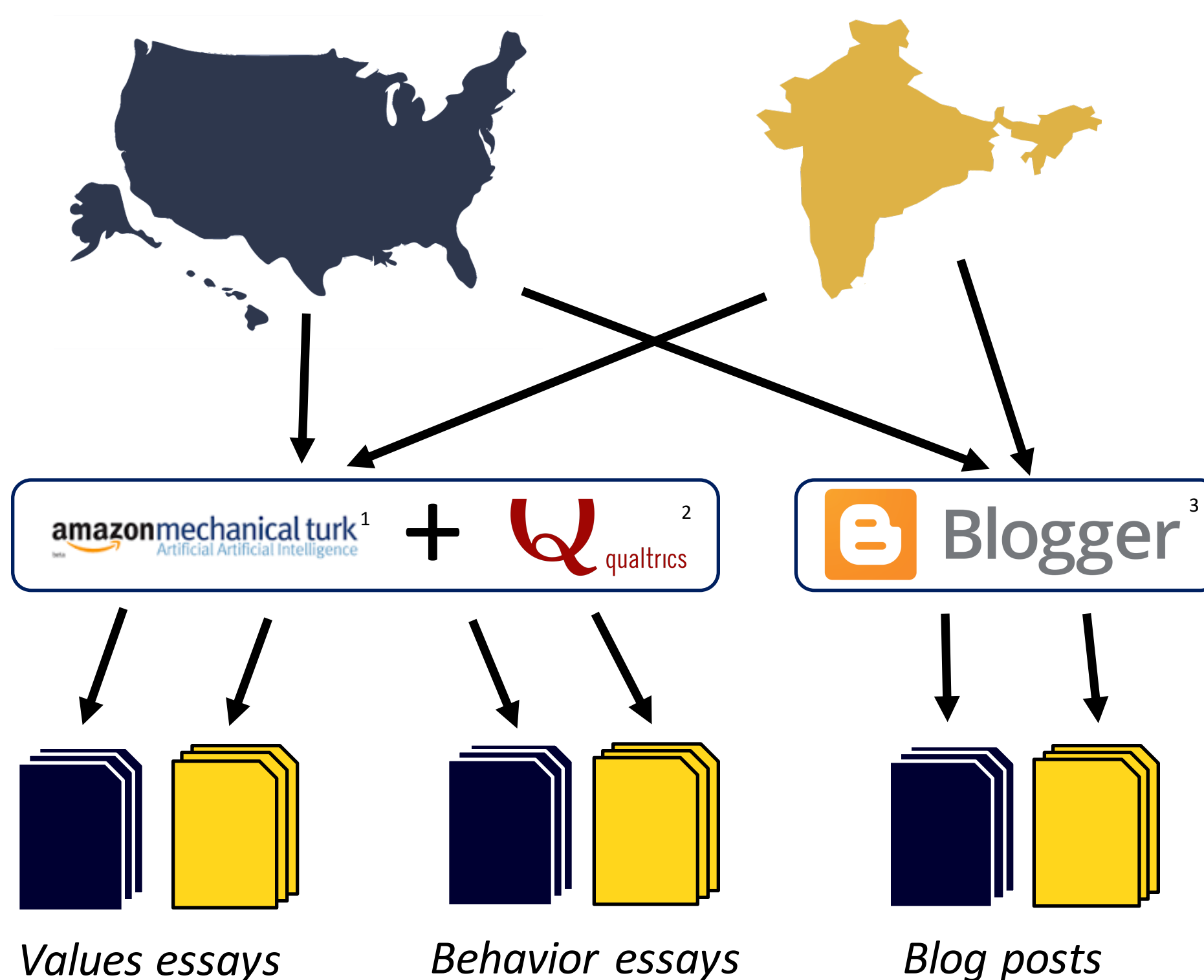


Introduction

- What are personal values? (Schwartz and Bilsky 1990)
 - Concepts or beliefs
 - About desirable states or goals
 - Transcend specific situations
 - Guide selection of evaluation of behavior
 - Are ordered by relative importance
- How can we measure values?
 - Rokeach values survey (Rokeach 1988)
 - Schwartz values survey (Schwartz 1992)
 - World values survey (Inglehart et al. 2005)
 - Topic model analysis** on open-ended text surveys (Boyd et al. 2015)
- What are our goals?
 - Collect and analyze a **multicultural dataset** that captures both **values** and **behaviors**.
 - Determine how the **culture** of an author is **related** to the use of **value-specific topics**.
 - Explore **correlations** between the usage of value-related themes and behavior-related themes across multiple cultures.

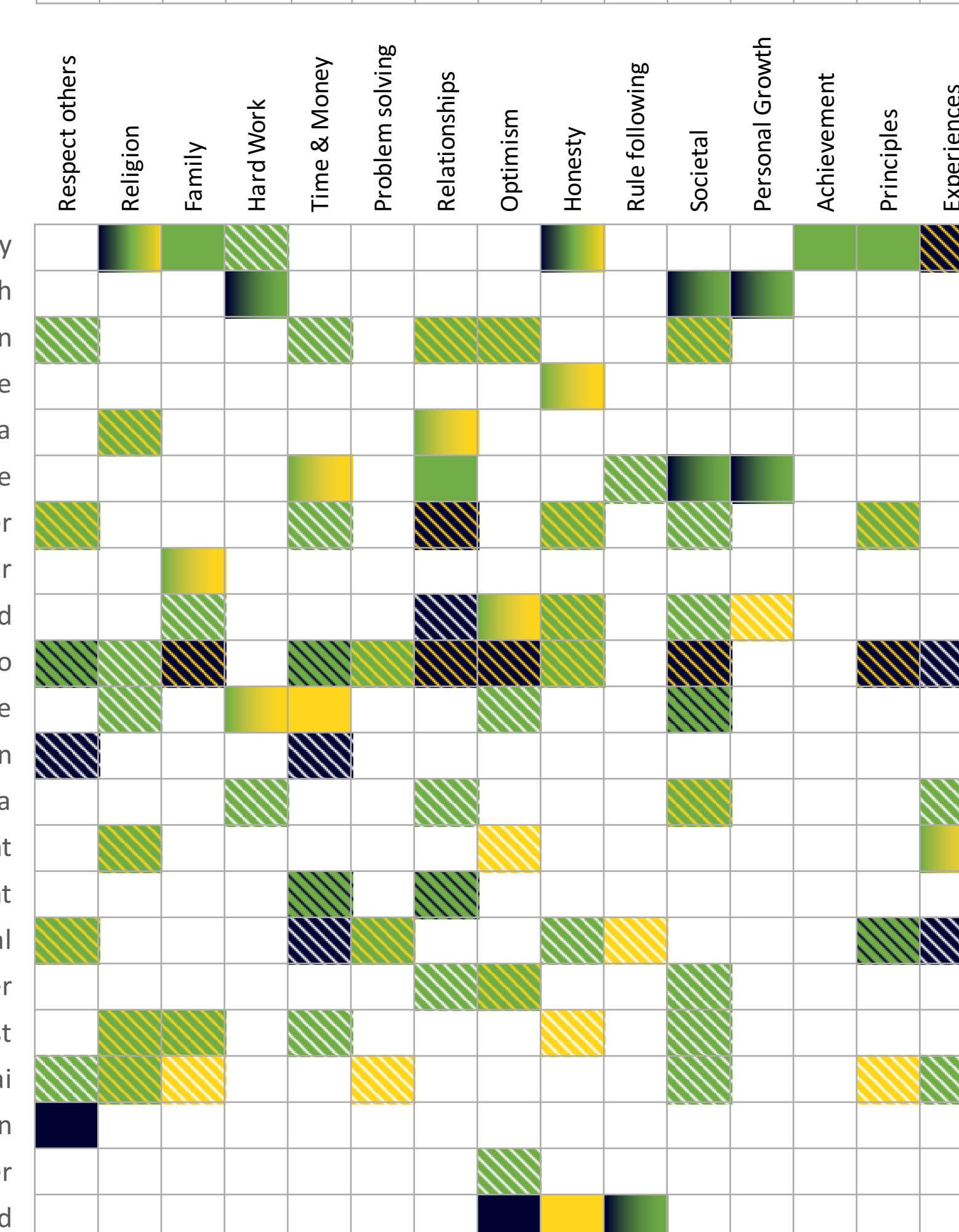
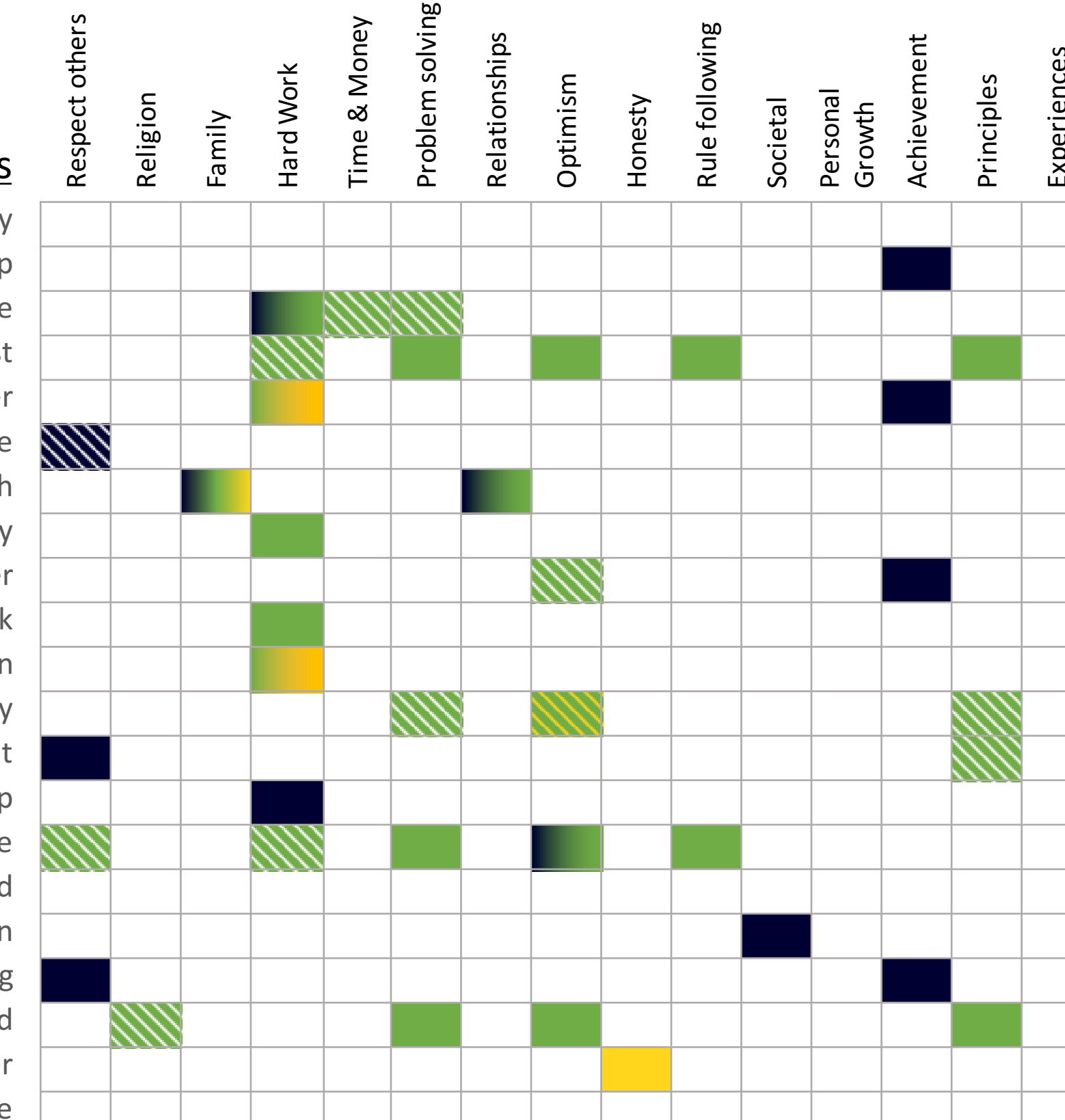
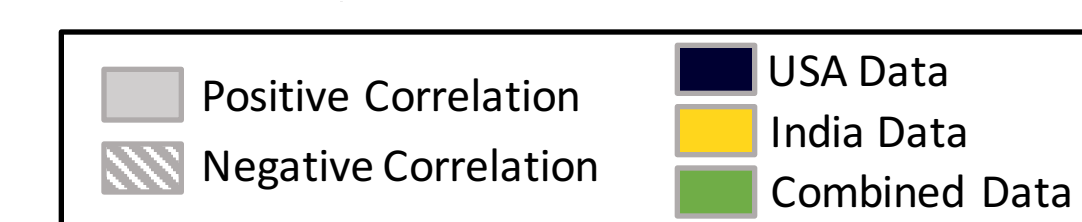
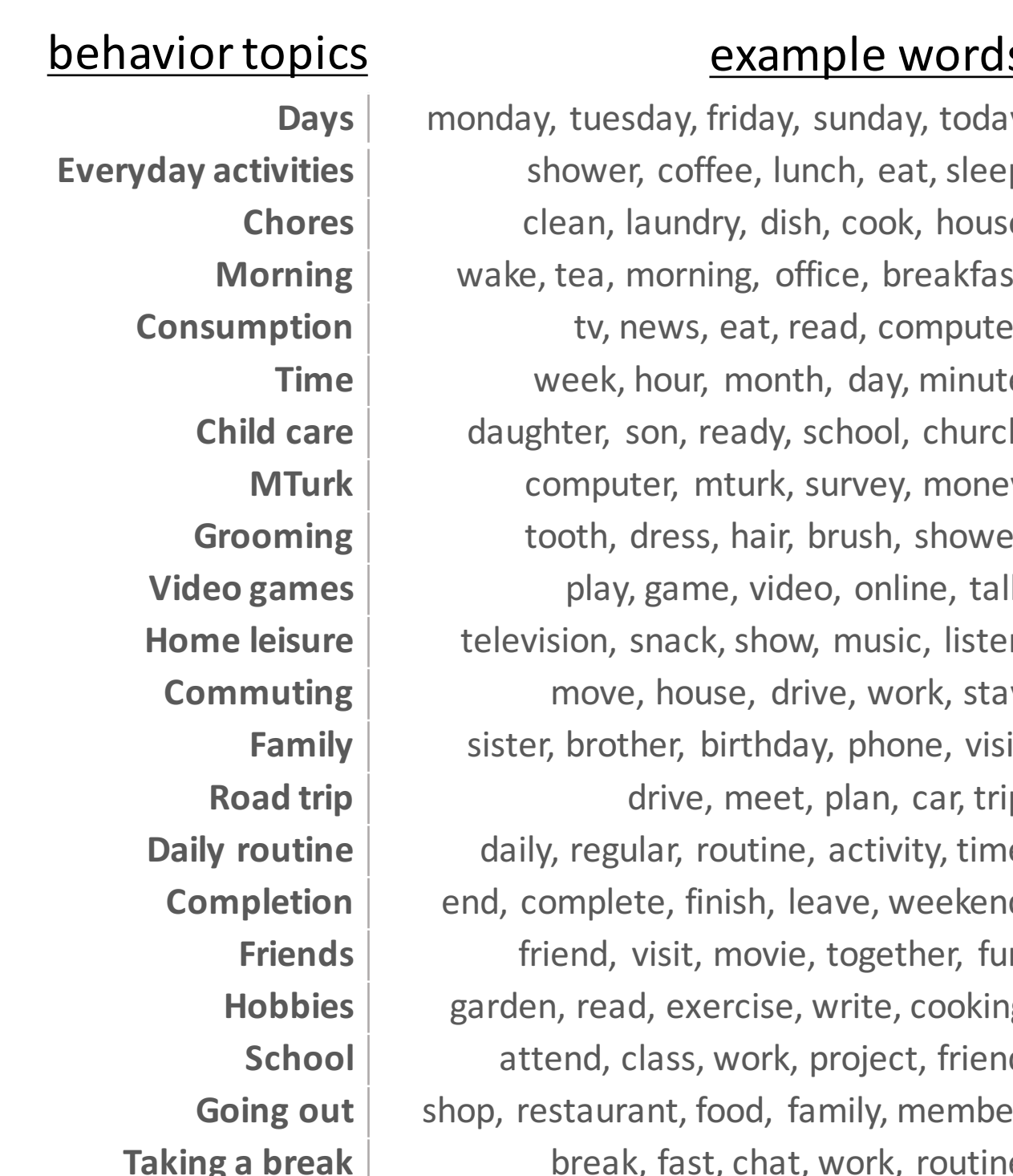
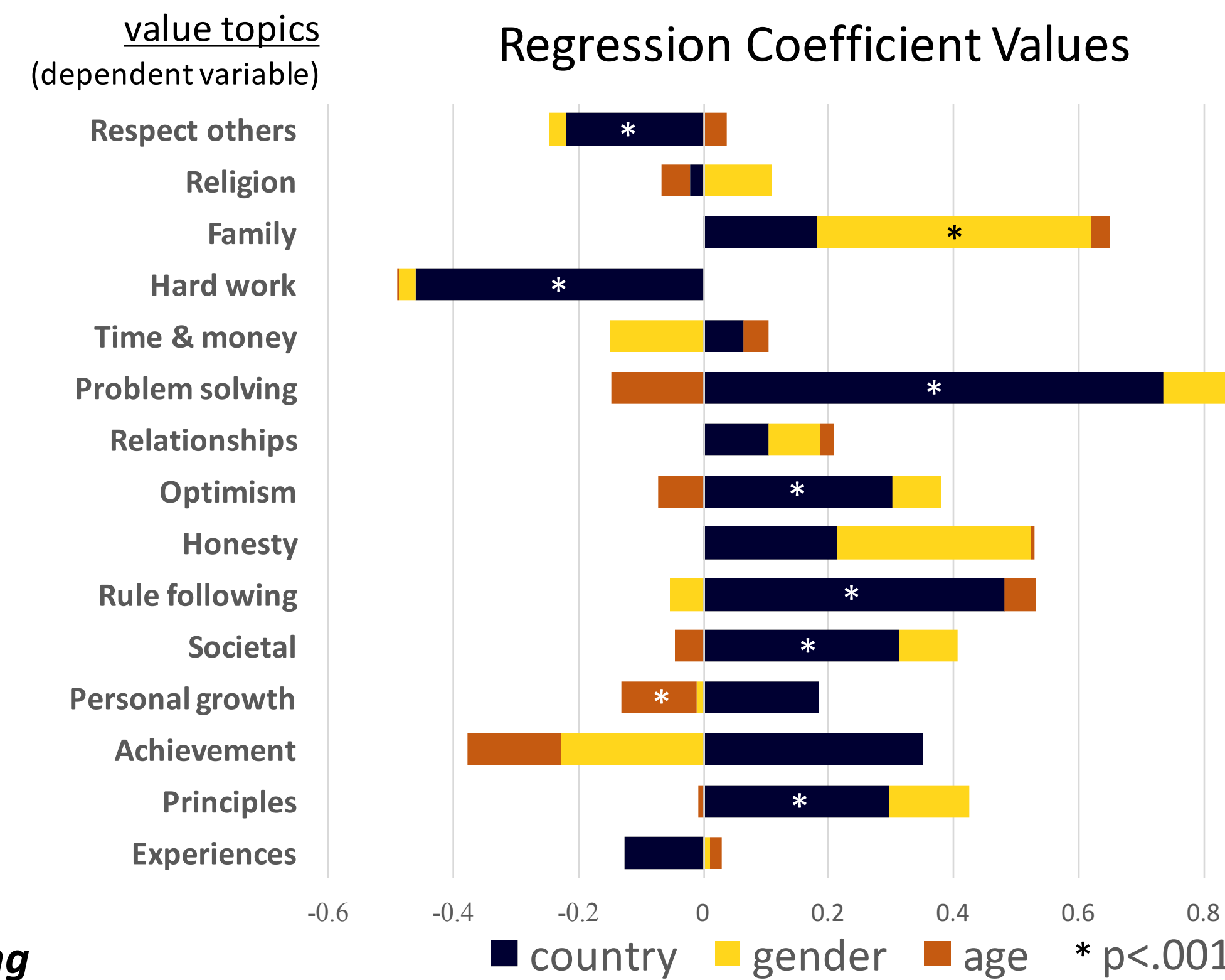
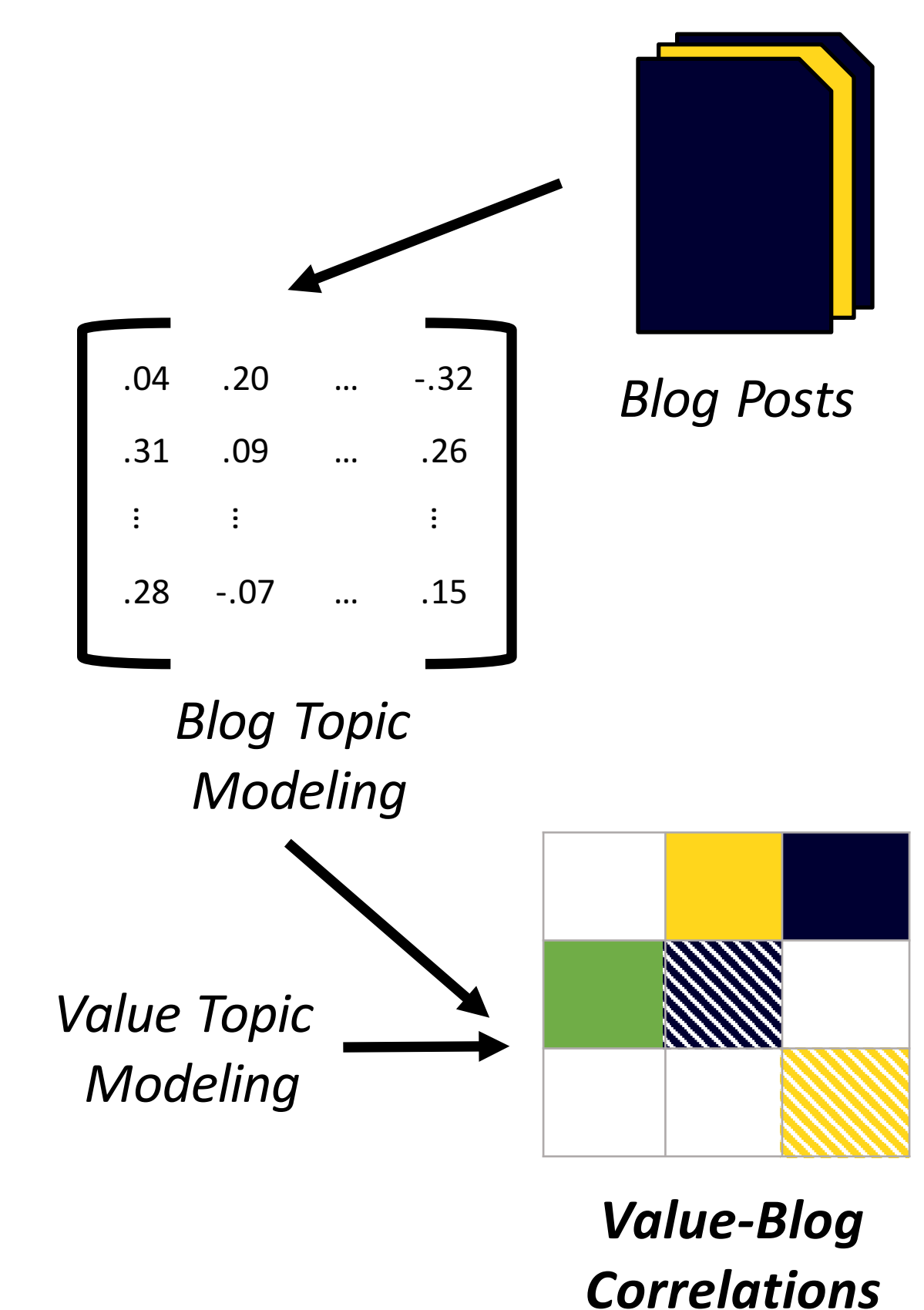
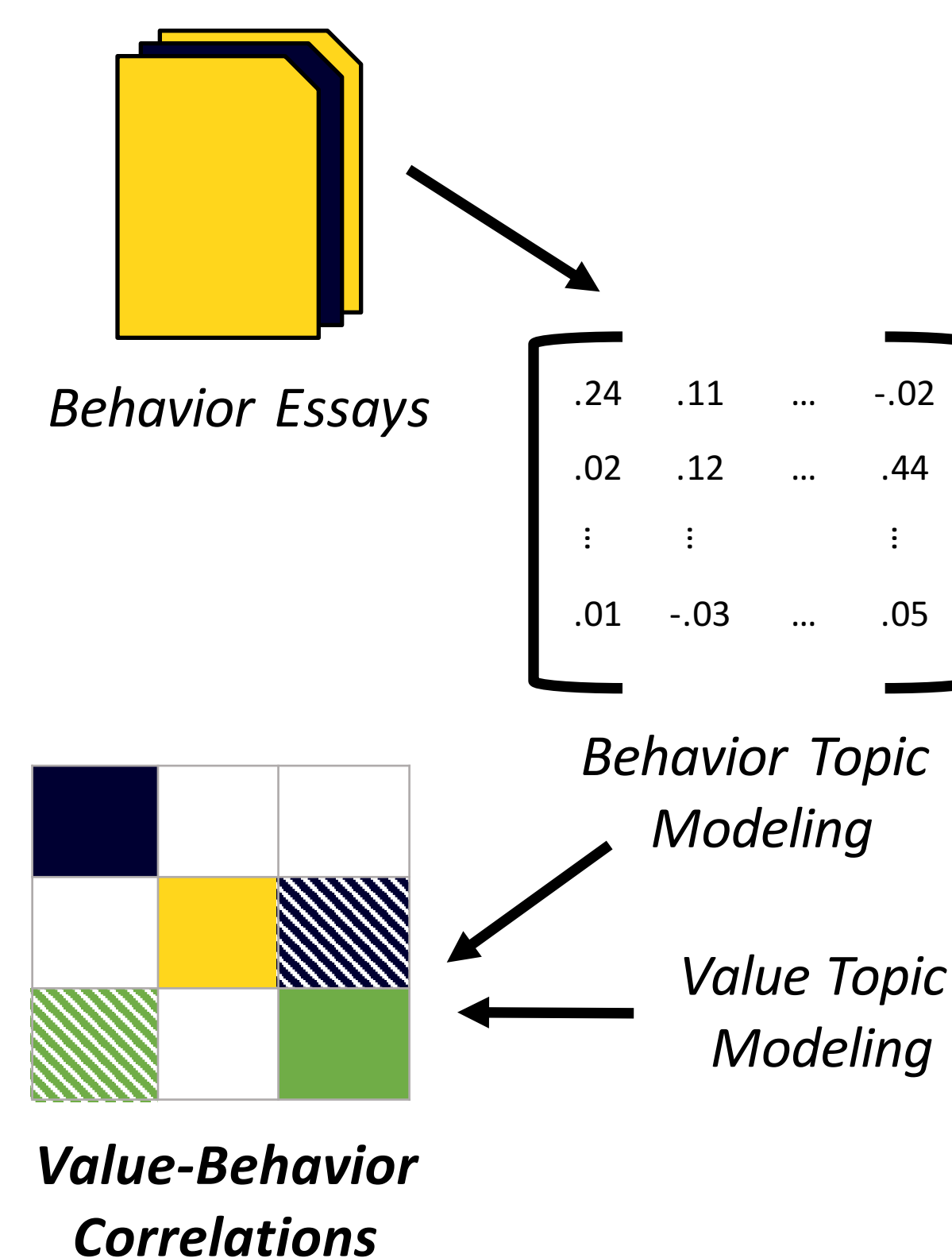
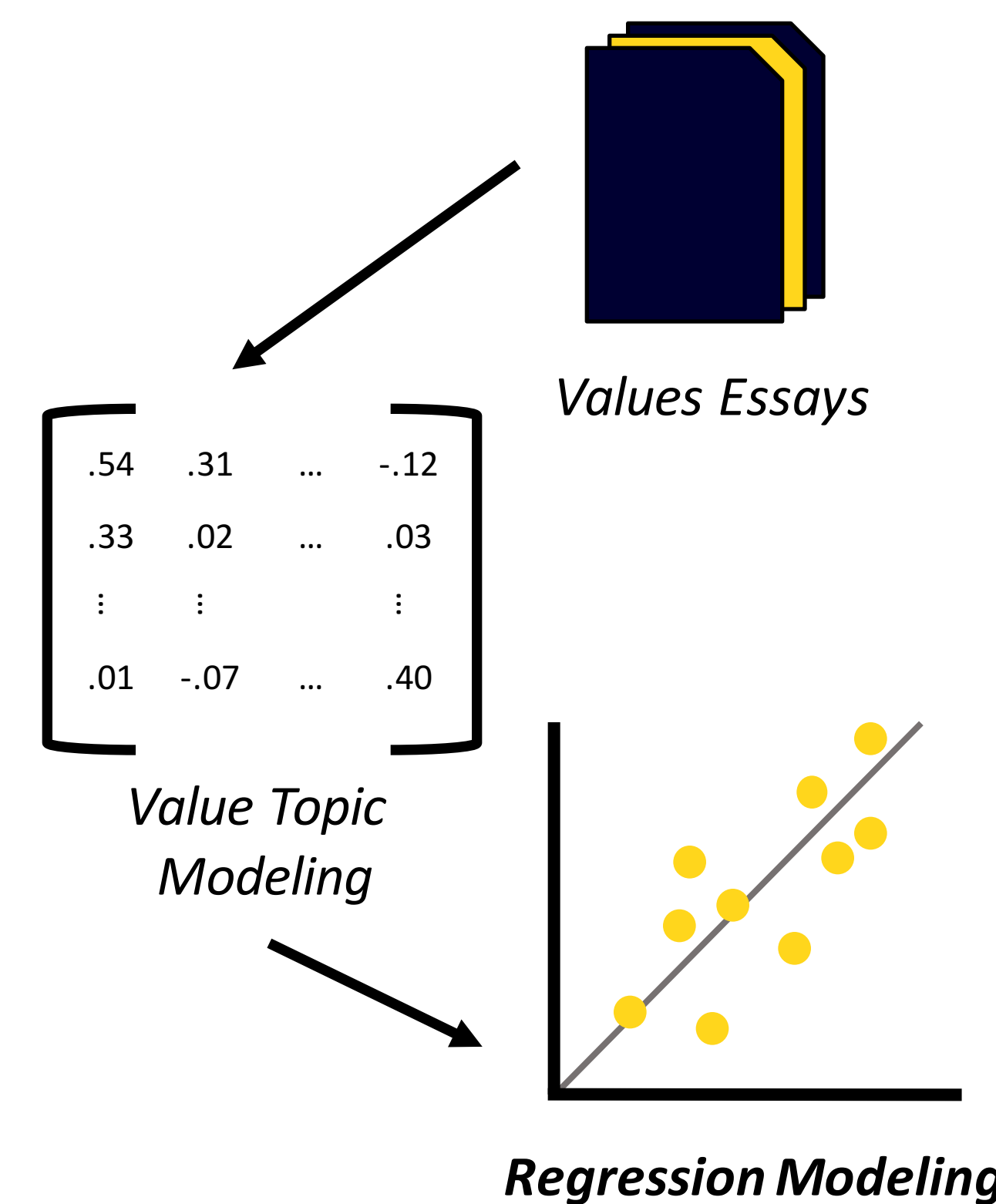
Data Collection

- We asked 1K people from the US and India about their most important values and a description of their recent activities.
- We downloaded 1K blogs for each country for every month between Jan 2010 and September 2015.



¹ mturk.com
² qualtrics.com
³ blogger.com

Method and Results



Regression Models

- Regress for each topic t :
$$usage(t)_i \sim \beta_0 + \beta_1 C_i + \beta_2 G_i + \beta_3 A_i + \epsilon$$
- Where:
$$usage(t)_i = 100 \times \frac{Usage\ of\ topic\ t\ by\ person\ i}{Average\ usage\ of\ topic\ t}$$
- And:
$$C_i = \text{country} \quad G_i = \text{gender} \quad A_i = \text{age}$$

Conclusions

- We explore cultural differences in the way people talk about personal values, and how values-related topics correlate with behavior-related topics.
- We use a regression model to disentangle the influence of several covariates on the usage of a particular topic by a given author.
- Using personal values as an example domain, we present a generalizable methodology that could be applied to various open-ended survey response data.
- Future work should explore the effect of changing the underlying topic model used and its parameters, and seek to also disentangle the influence of covariates on the inter-topic correlations.

References

- Boyd, R. L., 2015. MEH: Meaning Extraction Helper (Version 1.4.05) [Software] Available from <http://meh.ryanb.cc>.
- Boyd, R. L.; Wilson, S. R.; Pennebaker, J. W.; Kosinski, M.; Stillwell, D. J.; and Mihalcea, R. 2015. Values in words: Using language to evaluate and understand personal values. In *Ninth International AAAI Conference on Web and Social Media*.
- Chung, C. K., and Pennebaker, J. W. 2008. Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality* 42:96–132.
- Inglehart, R., Puranen, B., Pettersson, T., Nicolas, J.D. and Esmer, Y., 2005. The World Values Survey.
- Rokeach, M., 1988. *Rokeach value survey*. Consulting Psychologists Press.
- Schwartz, S. H., 1992. Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology*, 25:1-65.
- Schwartz, S. H. and Bilsky, W. 1990. Toward a psychological structure of human values. *J. Personality and Social Psychology*. 58:878-891

Acknowledgements

This material is based in part upon work supported by the National Science Foundation (#1344257), the John Templeton Foundation (#48503), the Michigan Institute for Data Science, and the National Institute of Health (#5R01GM112697-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, the Michigan Institute for Data Science, or the National Institute of Health. Finally, we would like to thank Konstantinos Pappas for providing the code used to collect the blog data.