FACULTY RESEARCH AWARD APPLICATION – 2022-2023

Co-Investigator

First Name: Joseph
Last Name: Henrich
Email: henrich@fas.harvard.edu
Department: Human Evolutionary Biology
School: FAS
Grant Manager Name:
Grant Manager Email:

Co-Investigator

First Name: Anke
Last Name: Becker
Email: abecker@hbs.harvard.edu
Department: Harvard Business School
School: Harvard Business School
Grant Manager Name:
Grant Manager Email:

Co-Investigator

First Name: Max
Last Name: Winkler
Email: mwinkler@fas.harvard.edu
Department: Human Evolutionary Biology
School: FAS
Grant Manager Name:
Grant Manager Email:

Research Project

Title: Psychological Change: Evidence from 300 Years of US Local Newspapers
Start Date: June 1ˢᵗ, 2022
End Date: May 30ᵗʰ, 2022
Budget Requested from MBB: $50,000
Are other funds available for this project?: X Yes _____No
If yes, what source(s)? What is the total amount available from other sources? $ 10,000

Please check the applicable box or boxes. This is an application for support for:

☐ Postdoctoral Fellow
☐ Visiting Research Fellow
X Research Assistant
☐ Research Equipment or Supplies
☐ Subject Fees
X Interdisciplinary Workshops, Conferences, or Seminars
X Other - Please explain in a few sentences: Part of the funds will be used to purchase a corpus of historical US newspapers containing more than 250 million pages, covering all 50 states, and spanning the past 300 years. The seller is newspaperarchive.com, one of the largest providers of

historical newspapers. This will be the largest newspaper corpus available to academics to date, offering us an unprecedented opportunity to study the evolution of culture and psychology through the lens of historical text.

As you complete the following sections of this application

1. Please know that your proposal will be reviewed by a committee whose members are drawn from a variety of disciplines. Please avoid specialized discipline-specific vocabulary whenever possible. Where such terminology or language is felt to be required, please define usage in language appropriate to a general academic readership.
2. Please adhere to page (or word number) limits and please use a standard font type and size wherever and whenever possible.
3. Please offer a clear account of the role to be played by personnel for whom salary is requested, be it technical or postdoc. In the case of postdoc, the role should be substantial with respect to the intent of the program and not simply the task of organizing symposia, etc. And please include a budget justification.
4. Please remember that funding can be awarded only for a single year at a time.

If you are applying for support for a fellowship*, a research or support staff member*, or research expenses, please complete this section.

A. **Technical Description (no more than 300 words)**
Please provide a succinct description of the problem, background theory, relevant hypotheses and planned methods.

Recent research documents substantial psychological variation around the globe. Yet, our understanding of the cultural evolutionary forces that shape this psychological variation is severely hampered by the lack of longitudinal psychological data. At the heart of this research project is the creation of a novel US-wide dataset that captures regionally fine-grained psychological variation stretching back more than three centuries. This dataset will allow us to address three interrelated questions on psychological change: First, how do aspects of psychology such as individualism, moral universalism, and normative tightness evolve across time and space in the United States? Second, how is this psychology shaped by underlying factors such as migration, social structure, or natural disasters? Third, how does this psychology coevolve with economic and political institutions over time? To address these questions, we aim to acquire a large, geo and time-tagged corpus of newspapers published in the United States over the last three centuries. By relying on state-of-the-art methods from computational linguistics (e.g., contextualized language models), and informed by theories from cultural evolution, we will extract aspects of psychology from newspapers quantifying cultural change at an unprecedent

level of granularity both in terms of temporal and spatial resolution. Our proposed effort is theoretically novel (integrating multiple perspectives that have been advanced in isolation in individual fields) and methodologically innovative (by bridging subject expertise from fields that are traditionally considered unrelated). Notably, the proposed research is not solely of hypothesis-testing nature: while we do test specific hypothesis as outlined in the Project Description, we also (a) make available a large corpus for future research and make methodological advances on psychology from text integrating insights from cultural evolution, computational linguistics, economics, and psychology. Overall, the proposed research will advance our understanding of cultural change in the United States over the course of several centuries.

B. Non-Technical Description (no more than 300 words)
Using non-technical terms, please describe the set of questions that your project aims to address and the methods you will use to answer them. Describe how this work fits with the mandates of MBB, focusing especially on its interdisciplinary nature.

As our understanding of culture becomes more nuanced across disciplines, cross-disciplinary efforts become crucial in understanding how, why, and when culture changes. Here, we propose a research project that integrates methods and insights from psychology, economics, anthropology, and computational linguistics to investigate cultural change across time and space in the US. A key characteristic of the research project is that we will create a data set on aspects of psychology that stretches back over several centuries. While (most) people who lived over the last three centuries are dead and we cannot give them psychological questionnaires about their values and political preferences or ask them to participate in an economic game, we seek to gain access to large corpora of text (e.g., newspapers). We will conduct a textual analysis in which we extract information on people's psychology from this large newspaper corpora (250 million pages from over 11,500 newspaper titles, in different locations). This allows us to create psychological indicators going back in time about 300 years. Based on those indicators, we will document regional and temporal variation in aspects of psychology, investigate how underlying factors such as migration, social structure, or natural disasters impact this psychology, and explore the co-evolution of this psychology with political and economic outcomes. We therefore expect that this

project will not only generate insights on psychological change but that it will deliver key insights on highly policy relevant topics including political polarization, formation of racial attitudes, and regional variation in economic prosperity.

C. Detailed Project Description

Please include a more detailed description of your proposed project (approximately 3 to 5 pages), including background, aims and hypotheses, methods, data collection and analysis (where appropriate), any preliminary or anticipated results, how this research is relevant to MBB, and any other information that you would like the review committee to consider.

Traditionally, social scientists have studied culture by comparing societies or social groups within a given point in time. However, culture can change, sometimes even change rapidly (e.g., within a generation), with important effects on people's cognition, emotions, and behavior (Henrich, 2020; Schulz et al., 2019). A prominent example of rapid cultural change is the potential decline in social trust and civic mindedness in the United States in recent decades (Putnam 2000). Thus, understanding how, why, and when culture changes is an important scientific endeavor– in fact, understanding cultural change is considered as one of the biggest questions facing science according to the editors of *Science* (Kennedy and Norman 2005).

The interdisciplinary field of cultural evolution offers a theoretical framework for explaining cultural change (Henrich, 2016; Laland, 2017). This framework permits us to think systematically about the underlying factors and process that drives changes in norms, institutions, beliefs, and behavior. However, thus far, a systematic empirical analysis of many of the framework's central predictions has proven elusive due to the lack of high-quality, quantitative measures of cultural traits over time; typical tools such as surveys, behavioral experiments hardly stretch back in time.

This proposed interdisciplinary research aims to fill this gap by compiling a database of thousands of US local newspapers, covering all 50 US states and spanning the past 300 years, and by applying modern analytic methods such as Natural Language Processing (NLP) to obtain temporal measures of key psychological traits including individualism (Hofstede, 2003), moral universalism (Enke, 2019), normative tightness vs. looseness (Gelfand et al., 2011), and analytic thinking (Talhelm et al., 2014). Newspaper text can be informative about the cultural traits prevailing among the readerships, such as individualism, because newspaper editors and contributors often

respond to readers' goals and demands (Gentzkow and Shapiro 2011). Thus, the proposed methodology of measuring cultural traits from newspaper texts has the distinctive advantage of providing high-frequency and regionally fine-grained measures that stretch far back in time.

More specifically, the proposed research, drawing on a collection of diverse methodological perspectives and theoretical insights, aims to address three interrelated questions of interest to multiple fields including psychology, anthropology, economics, cultural evolution, and political science. These three questions are:

1. How did psychological traits, such as individualism, moral universalism, normative tightness, among others, evolve across US counties over the past 300 years?
2. How were these cultural traits affected by factors related to immigration, social structure, natural disasters (i.e., the "dust bowl"), and conflict (e.g., the Civil War)?
3. To what extent did these cultural traits coevolve with economic and political outcomes, such as economic productivity (urban growth), innovation (patents), and political behavior (voting)?

To empirically address these questions, we propose a program of research comprising four steps, each building on a cross-disciplinary collaboration among our team members. First, to provide the primary resource for our textual analysis analyses, we plan to use a significant part of the grant funds to purchase all issues of all US newspapers provided by newspaperarchive.com, one of the leading digital archives of historical and current newspapers. The provider of the archive scans newspapers and generates machine-readable text using optical character recognition (OCR). The database is one of the most comprehensive digital newspaper archives currently available: it contains more than 250 million pages from over 11,500 newspaper titles – ranging from big-city dailies to rural weeklies – and keeps growing every day. The database also provides newspaper metadata, including the date of publication and the place of publication for each newspaper. We have verified that this corpus has excellent coverage across space and time: It has at least one newspaper title in each US state and spans more than 300 years. The owner of newspaperarchive.com agreed to give us access to the text files of all newspaper pages of US newspaper titles for a one-time fee of $40,000. We will also be able to transfer the files to

Harvard servers and aim to make the database available to all interested researchers within the Harvard community.

Second, we will collect additional metadata from annual directories of US newspapers going back to 1869. Newspaper directories are standard sources for historical research on US newspapers but have only recently been digitized at large scale. They originated as a guide to potential advertisers and were intended to be complete. In each year, we will extract the name, city, circulation, selling price, and political affiliation of every newspaper and link these metadata to the text database purchased from newspaperarchive.com. Crucially, the circulation data will improve our understanding on where individual newspapers were circulated (and read), and the political affiliation data will enable us to examine highly specific hypotheses on the interplay between political institutions and cultural change in the US

Third, we will use the newspaper corpus to create text-based indicators that capture spatial and temporal changes in culture. Our text-based psychological indicators will rest on three different approaches:

1. We will build a diverse set of linguistic features indicative of individualist attitudes drawn from psychological language analyses. We will use standard features, such as pronoun ratios, together with frequencies of verb forms and topical diction to support a broad-spectrum analysis of psychological change. To measure individualism and self-concepts, for example, we will look at the relative frequency of the use of first-person singular vs. plural verbs. In conjunction with the 20 Statements Test ("I am _____"), we will also examine the frequency with which first-person singular expressions primarily convey a personal, or social relationship (Kitayama et al., 2009). For example, this allows us to examine whether, as Putnam argues, there has been a shift from individualist expressions to relational ones in recent decades.

2. Building on existing work that aims to measure moral values and norm tightness in textual data (Graham et al., 2009; Jackson et al., 2019), we will develop dictionaries of key words that are indicative of certain psychological inclinations or tendencies in historical newspapers. Graham et al. (2009), for example, developed the Moral Foundations Dictionary which captures aspects of morality based on a set of keywords. We will develop

an expanded dictionary for the newspaper corpus using novel collocation techniques (Kennedy et al., 2019).

3. We will extend existing work and explore novel ways to draw measures of social trust and individualism using the most recent advances from Natural Language Processing (NLP), such as GloVe (Pennington et al. 2014) and BERT (Devlin et al. 2019), two state-of-the art techniques that construct high-dimensional vector representations of text.

Fourth, we anticipate the following results:

1. A descriptive set of visualizations that demonstrate the evolution of the set of key cultural traits since the 18th century and the heterogeneity in this evolution, e.g., across different US regions, urban versus rural counties, or highly innovative counties vs. the rest of the country.

2. Causal analyses on the underlying factors that drive cultural change. In particular, we are interested to learn how immigration, social structure, natural disasters and conflict impacts psychology. We have already complied much of the necessary data on those underlying factors (Winkler et al., in prep.; Winkler, 2022). For example, we will merge key indicators of migration and social structure (based on the full count US census stretching all the way back to the 18th century) with our novel newspaper-based psychological indicators. To provide causal estimates, we will rely on quasi-random exogenous shocks (e.g., natural disasters) or use econometric techniques such as Instrumental Variables (IV) estimation. For example, to explore the role of immigration and social structure we will rely on IV-approaches of the immigration literature (Burchardi et al., 2019).

3. We will investigate the coevolution and downstream effects of the changing psychology on economic and political outcomes. For example, we'll ask whether an individualistic-impersonal psychology impacts or co-evolves with economic productivity (proxied by urban growth, or occupational diversity), innovation (patents), and political behavior (voting). Again, the historical US full-count Census will be a major data source, but we will also rely on other sources such as voting records. Importantly, the acquisition of the newspaper data set will not only benefit us but the Harvard research community more

generally. The psychological traits we extract in this project are only the tip of the iceberg of the information that can be obtained from the text of historical US newspapers. We aim to provide the Harvard community with access to the text data underlying our results, so that other researchers at Harvard can explore the corpus and conduct their independent research.

Finally, we request $5,000 to pay for research assistance by Olivia Zhang, a Harvard undergraduate in Applied Mathematics and specialized in NLP. We know already well from her help on prior projects and have the highest confidence that she will provide invaluable help in managing the newspaper database and analyzing the text data.

The present program of research is highly interdisciplinary, adopting methodological tools from multiple fields including computational linguistics, psychology, anthropology, cultural evolution, and economics. For example, we will draw on computational linguistics and psychology to create text-based indicators of aspects of psychology; we will rely on econometric tools of economists to estimate how different underlying factors may causally impact aspects of psychology; and we will incorporate theoretical insights from all listed disciplines, particularly cultural evolution, to inform our hypothesis.

Our results will be at the forefront of the emerging interdisciplinary research area that investigates the determinants and consequences of psychological variation using state-of-the-art computational methods only made available a few years ago. This project has the potential to become a flag-ship study since we are not aware of any study that leverages psychological change at such fine regional granularity while stretching back several centuries in time. Our highly interdisciplinary team is uniquely well-equipped to address our principal research questions using the above-mentioned methods:

Dr. Joseph Henrich is a professor at the Department of Human Evolutionary Biology and focuses on evolutionary approaches to psychology, decision-making and culture, and includes topics related to cultural learning, cultural evolution, culture-gene coevolution, human sociality, prestige, leadership, large-scale cooperation, religion, and the emergence of complex human institutions. He integrates ethnographic tools from anthropology with experimental techniques

drawn from psychology and economics. Dr. Anke Becker is an Assistant Professor in the Entrepreneurial Management Unit at Harvard Business School, specialized in economic development, political economy, and economics of gender. Her recent work examines the evolution of culture and gender, specifically gender-specific norms and how they help us understand entrepreneurship in developing countries. Dr. Max Winkler is an economist with expertise in development, culture, economic history, and political economy. He studies the determinants of cultural change and how social and cultural factors drive long-run economic development. His recent work pioneers the use of newspaper articles to extract psychology. Furthermore, his work examines the origins and effects of cultural tightness, beliefs in supernatural protection, racism in media, and socio-cognitive diversity. Dr. Joseph Dexter is a computational biologist by training with interests in applied data science. His research is concentrated in two interdisciplinary areas: digital humanities including computational text analysis for Latin, ancient Greek, and other premodern traditions and the cultural evolution of literature, and systems biology and mathematical modeling for biomedicine. Dr. Mohammad Atari is a social psychologist, interested in why and how morality promotes cooperation and conflict. He is currently examining the psychological and evolutionary processes that underlie cultural change and development of institutions using a collection of methodological approaches, including social psychological experiments, survey-based assessment, computational modeling, and historical text analysis. Dr. Jonathan Schulz is an assistant professor of economics at George Mason University. His research lies at the intersection of economic history, experimental economics, and cultural evolution. His interdisciplinary studies on cultural and psychological variation have been published in the leading general interest journals.

References

Burchardi, K. B., Chaney, T., & Hassan, T. A. (2019). Migrants, ancestors, and foreign investments. Review of Economic Studies, 86(4), 1448–1486.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv:1810.04805 [Cs].

Enke, B. (2019). Kinship, Cooperation, and the Evolution of Moral Systems. The Quarterly Journal of Economics, 134(2), 953–1019.

Gelfand, M.J., Raver, J.L., Nishii, L., Leslie, L.M., Lun, J., Lim, B.C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I.C., Fischer, R., Fülöp, M., Georgas, J., Kashima, E.S., Kashima, Y., Kim, K., Lempereur, A., Marquez, P., Othman, R., Overlaet, B., Panagiotopoulou, P., Peltzer, K., Perez-Florizno, L.R., Ponomarenko, L., Realo, A., Schei, V., Schmitt, M., Smith, P.B., Soomro, N., Szabo, E., Taveesin, N., Toyama, M., Van De Vliert, E., Vohra, N., Ward, C., Yamaguchi, S., 2011. Differences between tight and loose cultures: A 33-nation study. Science (80-.). 332, 1100–1104.

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), 35–71.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046.

Henrich, J. (2016). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smart. Princeton University Press.

Henrich, J. (2020). The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous. Farrar, Straus and Giroux.

Hofstede, G., 2003. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations. SAGE Publications.

Jackson, J.C., Gelfand, M., De, S., Fox, A., 2019. The loosening of American culture over 200 years is associated with a creativity–order trade-off. Nat. Hum. Behav. 3, 244–250.

Kennedy, B., Atari, M., Davani, A. M., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral concerns are differentially observable in language. *Cognition*, *212*, 104696.

Kennedy D, Norman C. What don't we know? Science. 2005 Jul 1;309(5731):75.

Kitayama, S., Park, H., Sevincer, A.T., Karasawa, M., Uskul, A.K., 2009. A cultural task analysis of implicit independence: comparing North America, Western Europe, and East Asia. J. Pers. Soc. Psychol. 97, 236–55.

Laland, K. N. (2017). Darwin's Unfinished Symphony: How Culture Made the Human Mind. Princeton University Press.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Putnam, R. D. (2000). Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. Science, 366(6466).

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., Kitayama, S., 2014. Large-scale psychological differences within China explained by rice versus wheat agriculture. Science 344, 603–8

Winkler, M., (in preparation). Cultural Change: Evidence from 300 Years of U.S. Local Newspapers.

Winkler, M., (2022). Do Disasters Affect the Tightness of Social Norms?

* If you are applying for support for a fellowship, please also complete this section.

Fellowship Candidate Information
Name:
Email address:
Degree: Degree date:
Current advisor:
Identify support the candidate is applying for that would run concurrently with the period covered by this application (type, source, amount, dates):
Identify the candidate's other potential funding sources for this project:
Please include the candidate's CV with this application packet.

If you are applying for support for developing and launching a workshop, conference, seminar, symposia, course, or other non-experimental project, please provide the following:

A. **Introductory summary/hypothesis**
B. **Background information, including any preliminary work for this project**
C. **Significance, expectations, possibilities, alternative perspectives**
   Please discuss a projection of themes likely to arise from the effort that will be carried forward jointly within the two or more disciplines. When possible, comment on whether a potential cadre of interested faculty or students has been identified, whether the subject or topic matter is substantially mature with cogent texts/sources already available, and whether a set of likely and plausible speakers have been identified/approached.
D. **Statement of relevance for MBB (Why MBB should support the project)**
E. **Progress report if this is a request for a second year of funding**

Application Packet Checklist
☐ Application form
☐ A budget (next page)
☐ Applicable supporting documents

Please email a single pdf version of this application packet to HarvardMBB@gmail.com by March 21, 2022.

Mind Brain Behavior Interfaculty Initiative Faculty Award
2022-2023 BUDGET PAGE
PI(s): Anke Becker, Joseph Henrich, Max Winkler
Project Title: Psychological Change: Evidence from 300 Years of US Local Newspapers

Expense Category	Budget Amount	Description
Personnel	**$ 5,000**	**RA Olivia Zhang**
Other services		
Supplies & Equipment		
Human or animal subject fees		
Printing/postage		
Travel:	**$ 5,000**	**Dissemination: Traveling to conferences**
Workshop or Conference Expenses		
Other: Newspaper data	**$40,000**	**Purchase of a corpus of historical US local newspapers**
Budget Total	**$50,000**	