[In press, *Current Directions in Psychological Science*]

Historical Psychology

Mohammad Atari

Joseph Henrich

Department of Human Evolutionary Biology, Harvard University

Word Count = 3,397 words

References = 40

Author Note

Correspondence concerning this article should be addressed to Joseph Henrich, Department of Human Evolutionary Biology, Harvard University, 11 Divinity Ave, Cambridge, MA 02138. E-mail: henrich@fas.harvard.edu

Abstract

A growing body of evidence suggests that many aspects of psychology have evolved culturally over historical time. A combination of approaches, including experimental data collected over the last 75 years, cross-cultural comparisons and studies of immigrants, points to systematic changes in psychological domains as diverse as conformity, attention, emotion, morality and olfaction. However, these approaches can only go back in time for a few decades and typically fail to provide continuous measures of cultural change, posing a challenge for testing deeper historical psychological processes. To tackle this challenge most directly, computational methods emerging from natural language processing can be adapted to extract psychological information from large-scale historical corpora. Here, we first review the benefits of psychology as a historical science, and then present three useful classes of text-analytic techniques for historical psychological inquiry: dictionary-based methods, distributed-representational methods, and human-annotation-based methods. These represent an excellent suite of methodologies that can be used to examine the record of "dead minds." Finally, we discuss the importance of going beyond English-centric text analysis in historical psychology to foster a more generalizable and inclusive science of human behavior. We propose that historical psychology should incorporate and further develop a variety of text-analytic approaches to reliably quantify the historical processes that gave rise to contemporary social, political, and psychological phenomena.

Keywords: historical psychology, natural language processing, cultural evolution, culture.

Historical Psychology

To many psychologists, who may implicitly rely on a digital-computer metaphor of the mind, studying history may seem peculiar, if not irrelevant. However, many evolutionary researchers now argue that we are a "cultural species" whose brains evolved genetically to *ontogenetically* (i.e., during development) acquire and ingrain culturally-specific ways of processing information (Henrich 2016): our thoughts, emotions, and behaviors are shaped by (and shape) our societies, and our societies are shaped by (and shape) their histories (Henrich, 2020; Markus & Hamedani, 2020; Uchiyama et al., 2022). Historical contexts do not exist apart from people: institutions (Schulz et al., 2019), technologies (Frank & Barner, 2012), wars (Henrich et al., 2019), and ecological disasters all (Vardy & Atkinson, 2019) shape our minds and our mind then shapes history in a sequence of interdependent actions that reflect and reinforce each other. Hence, a fuller understanding of contemporary human psychology requires understanding the historical contexts that led to our present-day psychology.

To put "historical psychology" itself into a historical context, a number of psychologists have proposed to integrate cultural history and psychology. In the early 1930s, Vygotsky and colleagues argued that understanding psychological processes required considering four different perspectives: phylogeny, cultural history, ontogeny, and microgenesis (Vygotsky & Luria, 1930/1993). Cole (1990), who traces his thinking in cultural-historical psychology to Soviet psychologists Lev Vygotsky, Alexander Luria, and Aleksei Leontiev, focused on cultural contexts as defined by a continual flow of constructed activity. Cole (1990) suggests that humans enter a world that is transformed by "the accumulated artifacts of previous generations." Therefore, culture becomes history in the present, and our social psychology is the study of human behavior in contemporary history (Gergen, 1973). Although these early cultural–

historical psychologists are in some respects out of date, their emphasis on the role of culture and history in psychological processes and their inclusive evolutionary framework provide useful frameworks to build upon. Indeed, these efforts have been important steps in moving toward making psychology a genuinely universal and inclusive science.

Recently, equipped with the field of cultural evolution, with its integration of evolutionary theory, focus on adaptive learning and application of formal mathematical models, Muthukrishna et al. (2021) argued that for psychology to develop into a mature science of human behavior, it needs to develop into a historical science. In this paper, we define historical psychology as research that examines the ways in which histories and psychologies make each other up in a continuous dynamic of mutual co-constitution—a process termed cultural niche construction (see Ihara & Feldman, 2004). We argue that historical psychology holds the potential to deepen our understanding of human behavior, and when paired with evolutionarily informed theories and state-of-the-art methodologies in Natural Language Processing (NLP), it can provide widely relevant insights into psychological science.

Over the last half century, research in both cultural and cross-cultural psychology has documented substantial psychological variations across populations (e.g., Nisbett, 2003) —some of these studies have been referred to as “geographical psychology” (Rentfrow & Jokela, 2016); meanwhile, temporal variation in aspects of psychology within the same population over decades, centuries and even millennia represent the core of the nascent field of historical psychology. In other words, geographical psychology has been productive in beginning to map the contemporary spatial variation in psychology, but relatively little attention has been paid to “when” and “why” it varies over time in the same population (see Varnum & Grossmann, 2017).

Cultural evolution offers a synthetic theoretical framework for explaining psychological differences across both time and space (Boyd & Richerson, 1985; Henrich, 2016; Muthukrishna & Henrich, 2019). Cultural evolutionary theory is an extension of evolutionary theory that describes the cumulative process by which various norms, technologies, values, and behaviors are selectively transmitted and retained through social learning as well as our species evolved psychology. Indeed, our cognitive architecture allows us to acquire adaptive beliefs and behaviors over time. Hence, cultural evolution offers a framework for explaining not only cross-societal psychological differences, but also cross-temporal ones.

Although a common folk model of cultural evolution emphasizes creative, conscious innovation in which inventors "buckle down, rack their brains and come up with something new" (Pinker, 1997) many empirical studies suggest that most novel ideas are actually recombinations of older ideas, which often meet serendipitously, that accumulated gradually over time (Henrich, 2016). Moreover, much of individual creativity depends on a cultural toolkit of cognitive gadgets that sharpen our thinking and shape our causal model construction with a repertoire of mental tools, heuristics, beliefs, norms, and values bequeathed to us by earlier generations. Hence, human psychology is best understood to have been shaped by millions of years of genetic evolution, thousands of years of cultural evolution, and a short lifetime of individual experience; and yet, much of psychological science has focused on that short lifetime of experience. Psychology still overwhelmingly generalizes from present-day populations living in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (Henrich et al., 2010). The WEIRD people problem is both a matter of geography and of history (Gray et al., 2010).

By taking historical context seriously, researchers have recently examined the historical origins of WEIRD psychology. Tacking back and forth between historical evidence and

contemporary psychological data, Henrich (2020) traces the emergence of WEIRD psychology back through the emergence of impersonal markets during the Commercial Revolution and the proliferation for voluntary associations, including guilds, monastic orders, charter towns and universities during the High Middle Ages, to the transformation of the families by the Catholic Church. Supporting this, Schulz et al. (2019) link contemporary psychological variation across a broad range of domains, including individualism, tightness (i.e., the strength of social norms), conformity, moral values, and impersonal prosociality (i.e., cooperation with, fairness toward, and trusting of strangers and anonymous others), back to both kinship organization (e.g., cousin marriage and polygyny) and the spread of the medieval Catholic Church, within Europe and globally. The idea here is that by dismantling the dense kinship networks of pre-Christian Europe through its marriage prohibitions (e.g., cousin marriage and polygamy) and regulation of inheritance and post-marital residence patterns, the Church shifted people's psychology, increased residential mobility and open the door to new social organization. To test this hypothesis, these authors assembled historical, ethnographic, and psychological databases. By tracking the historical diffusion of the Church's regional centers—bishoprics—across Europe, they calculated the duration of exposure to the Church from roughly 500 to 1500 CE and used the resulting data to predict contemporary psychological variation across Europe and around the world on four psychological measures: individualism, conformity, impersonal fairness, and impersonal trust. These authors found the Western Church (i.e., the branch of Christianity that evolved into the Roman Catholic Church) to transform European kinship structures during the Middle Ages resulting in a shift toward a WEIRDer psychology.

Historical texts, art, and archeological sources serve as a kind of "psychological fossil record" (Muthukrishna et al., 2021) that open up an opportunity to access data from dead minds.

The depth of our historical analysis is only bounded by how deep data can reliably go back in time. Past behaviors, norms, values, and narratives lie buried in historical artifacts, which range from archeological remains to written texts. These treasures are not only important for understanding the roots of modern psychological patterns, but also represent an untapped way of studying global psychological diversity (Slingerland, 2014). The dead represent a remarkably varied subject pool in terms of cognitive and cultural phenomena, especially compared with the samples typically studied by psychologists.

Although our inability to experimentally manipulate or unobtrusively observe historical participants places some limits on what we can infer from these (potentially decontextualized) data, such traces of human thought can be a rich and informative source of descriptive information on past psychology (Jackson et al., 2022). Of course, research can test theories about the drivers of psychological change by looking at "natural experiments." Natural experiments arise when historical events or factors—weather shocks, policy changes and arbitrary political boundaries—create quasi-random variation akin to experimental treatments. Such approaches can be effectively paired with and complemented by experimental approaches that use controlled experimental manipulation to test the same theories (e.g., Atari et al., 2022).

Text Analysis in Historical Psychology

As humans developed larger-scale societies over the course of history, the ever-expanding body of cultural information that was passed to the next generations expanded, which may have contributed to the evolution of writing system to efficiently transmit large amounts of information; hence, the analysis of written sources is a particularly important methodological toolbox in historical psychology. Fortunately, a great number of computational techniques

developed in NLP can be used or adapted for use in historical text analysis (for a review on how language analysis can advance psychology, see Jackson et al., 2022).

Atari and Dehghani (2022, p. 208) argue that "instead of qualitative analyses of divine texts or historical inscriptions, psychologists are often interested in quantifying language to understand, describe, explain, or predict the psychological characteristics of the producer of that language." These authors review psychological text analysis in studying social norms and moral values and find three major categories of methods in psychological text analysis (see Table 1): (a) dictionary-based methods; (b) distributed-representational methods; and (c) human-annotation-based methods. All these methodological approaches to text analysis can be used to quantify psychological constructs of interest in the past.

Table 1

Text-Analytic Methods and Their Application in Historical Psychology

Methodology	Description	Application	Threats to Validity
1. Dictionary-based methods	Developing wordlists that represent a psychological construct and counting how frequently these terms appear in a document	Quantifying the prevalence of a set of terms in different time units	a. Cherry-picking words to arrive at favorable evidence b. Including polysemous words in dictionaries c. Disregarding semantic lexical change over time
2. Distributed-representations methods	Representing words in the form of a vector that encodes the meaning of the word such that the words that are closer in the vector space are expected to be similar in meaning. Accordingly, the geometric relationship between these vectors captures meaningful semantic relationships between the corresponding words	Identifying analogies and quantifying the semantic similarity between a text or word and a particular set of terms in a high-dimensional space	a. Disregarding how words' frequency affects their vector representation as well as distance to other words and shift in meaning over time b. Using biased datasets to train word embeddings on c. Using fixed (vs. diachronic) word embeddings to examine psychological change over time
3. Human-annotation-based methods	Manual annotation of written language as ground truth based on subject knowledge to be used for training a machine-learning algorithm. This method accounts for compositional and sentence-level constructs.	Automating the labeling of historical textual data with regards to a psychological construct of interest	a. Non-experts might mislabel historical phenomena b. Regarding inter-annotator disagreement in subjective annotations as mere noise c. Using biased present-day knowledge bases to code historical concepts

Dictionary-Based Methods

One popular and simple text-analytic method is to apply dictionaries (or wordlists) to track historical trends. By measuring shifts in word frequencies over time, one can detect changes in psychology (although changes in norms could potentially result in changes in the meaning of words associated with different psychological dimensions; see Snefjella et al., 2019; for an example of change in the nomological network of a concept in a matter of decades, see Choi et al., 2021). Greenfield (2013), for example, found that words associated with individualism (e.g., "self") have become more frequent over the last two centuries. More recently, Choi et al. (2022) developed a threat dictionary, a linguistic tool that measures threat levels from textual data, and demonstrated this dictionary's validity in relation to objective threats in recent American history such as violent conflicts and pathogen outbreaks. Using data from newspapers that span over a century, Choi et al. (2022) found changes in threats to be associated with tighter social norms, collectivistic values, higher approval of sitting presidents, lower stock prices, and less innovation. Similarly, Winkler (2022) applied a dictionary of tightness–looseness to a corpus of U.S. newspapers from different regions of the United States since 1840. This provides a nearly continuous measure of tightness–looseness that varies through time and space, a unique combination of geographical and historical psychologies. Winkler (2022) demonstrated a long-term decline in average tightness as well as substantial spatial variation within the country. Comparing only the tightness–looseness of individual newspapers over time and across states, Winkler (2022) showed that economic declines cause people to tighten up, and that a one-percent increase in unemployment resulted in a rise in tightness corresponding to 6% of a standard deviation in normative tightness. Winkler (2022) then linked

these historical psychological shifts to both greater parochial cooperation and more votes for Donald Trump in 2016.

Another example of dictionary-based text analysis in historical psychology is a study by Scheffer et al. (2021) where the authors analyzed language in English books from 1850 to 2019, showing that the use of words associated with rationality (e.g., “determine” and “analysis”) rose after 1850, while words representing human experiences (e.g., “feel” and “hope”) declined. This pattern of language usage reversed over the past decades, paralleled by a shift from a collectivistic to an individualistic focus as reflected by the ratio of singular (e.g., “I”, “she”) to plural pronouns (e.g., “we”, “they”). These authors conclude that over the past several decades, there has been a marked shift in public interest from the collective to the individual, and from rationality toward emotion. Using a similar text-analytic approach, Martins et al. (2020) tested the hypothesis that early modern revolutions may be the product of long-term psychological variation, from hierarchical and dominance-based interactions to democratic and trust-based relationships. These authors showed an increase in cooperation-related words over time relative to dominance-related words in England and France, making the case for the important role of historical psychological changes in explaining the rise of early modern democracies.

While dictionary-based methods have been widely adapted by psychologists, in part because of their high interpretability and ease of use, their limitations should be noted. For example, in some cases simple lexical frequency changes may not be clear indications of psychological change. For instance, Scheffer et al.’s (2021) finding about rational words is confounded with the words in the “rational” dictionary (e.g., “analysis”) being highly prevalent in formal writings such as academic texts (Table 1 summarizes threats to validity).

Distributed-Representational Methods

Dictionary-based methods have practical challenges that limit their validity (see Kennedy et al., 2022). Distributed representations provide an alternative to the word-counting methods, capturing the relationship between contextually related words or larger chunks of text, rather than comparing the frequencies of words in documents. Modern methods of generating distributed representations of words using vectors have proven to efficiently provide representations that have excellent semantic regularities (for a review, see Kennedy et al., 2022). The nearest neighbors of terms in the semantic space tend to be highly meaningful. With distributed representations (word embeddings), we can ask a number of questions, such as how likely two words (or wordlists) are to co-occur in large textual data. For example, Garg et al. (2018) demonstrated how the temporal dynamics of embeddings enables us to quantify changes in stereotypes and attitudes toward women and ethnic minorities over time. Garg et al. (2018) integrated word embeddings trained on a century of text with the U.S. Census to demonstrate that changes in the word embeddings track closely with demographic and occupational shifts over time. By examining semantic similarities between particular groups of words, these authors tracked societal shifts (e.g., the women's movement in the 1960s) and also showed how specific occupations became more closely associated with certain populations over time. For example, around 1910, the top adjectives associated with Chinese last names were largely negative, including "irresponsible" and "barbaric." However, some qualitatively different adjectives emerged around 1990, with the same Chinese last names being closer with terms such as "inhibited" and "haughty." Using the same logic and methodology, Charlesworth and colleagues (2021) demonstrated the lack of variation in bias: these authors showed that gender bias, quantified via word embeddings, exists across textual data produced at different times and even by different age groups, both in children and adults.

As with other approaches, word embeddings and similar methods have limitations. First, the assumptions implicit in such off-the-shelf approaches may not always be clear to applied researchers who use them for historical text analysis. For example, van Loon et al. (2022) found that word embeddings are biased by word frequencies. Their analyses revealed that in word embeddings, highly frequent words tend to have positive associations in semantic space. Another important issue is that in studying lexical semantic change across time (i.e., detecting shifts in the meaning and usage of words), diachronic word embeddings (i.e., time-sensitive numerical representations of words that track meaning through time) are needed. But developing diachronic word embeddings remains a hard task because historical corpora are scarce. As such, it is crucial for historical psychologists to compile historical corpora.

Human-Annotation-Based Methods

Manual human annotation is the oldest approach and provides the ground truth for training machine-learning algorithms. In this class of methods, researchers agree on a theoretical framework with which they code text for the construct of interest (e.g., individualism). Then, a number of annotators code textual data for the presence of relevant information. An implicit presupposition of this approach is that historical data includes complex and indirect information; thus, human judges can best capture nuances and complexities of written text produced in the past (rather than, for example, relying on an a priori wordlist). Finally, a supervised machine-learning model is trained on these annotations and will be able to automatically identify the construct of interest in new corpora (for a review, see Atari & Dehghani, 2022; Slingerland et al. 2020).

While manual annotation can serve as a useful method in historical text analysis, there are issues to consider. For example, while manual coders can leverage their experience relative

to blunt methods such as word-counting, annotators can be biased by their demographics, values, and personality traits. These individual differences in manual coders give rise to disagreements on labels. Notably, disagreement in annotation of textual data is not always noise, it might reflect genuine uncertainties about a historical event, or individual differences of the annotators (for a review on dealing with annotation disagreement in subjective tasks in NLP, see Davani et al., 2022). Given temporal variations in the meanings of terms and changes in (unwritten) norms, non-expert annotators of today may not accurately code terms in a different time in a way that reflects how the term was understood during the time period being studied. Studies that involve multiple cultures should ideally use annotators who understand the sociohistorical context under investigation. Such issues are akin to issues raised by ethnographers who typically invest time into understanding concepts from the perspective of the population being studied.

Benchmarking

Like all measures in psychology, text-based measures should be examined for their validity (see Table 2). Prior work highlights the importance of benchmarking in historical text analysis (see Choi et al., 2022; Garg et al., 2018; Winkler, 2022). Researchers should validate their data against temporal and geographic ground truth (e.g., survey-based data) to make sure that their text analysis is picking up real psychological signal rather than noise or merely linguistic shifts with no meaningful psychological underpinning. For example, a measure of threat should reflect real historical events such as wars, famines, and social disarray. Some surveys have been conducted for decades (e.g., the World Values Survey, European Social Survey) and some online researcher-led platforms can offer valuable data (e.g., YourMorals.org, ProjectImplicit.net) that can be used to benchmark data extracted from written sources.

Table 2

Different Types of Validity for Text-Based Measures in Historical Psychology

Type of Validity	Definition	Historical Example
Face validity	The extent to which a text-based measure (e.g., a dictionary) appears to be appropriate for measuring a construct based on existing theories.	Text-based measures of collectivism in historical contexts should capture references to collectives (e.g., "group").
Convergent validity	The extent to which text-based measures (e.g., manual annotations) exhibit a strong relationship with scores on conceptually similar tests or instruments (e.g., surveys).	Text-based measures of threat should align with real-world threats in history (e.g., wars, ecological disasters).
Predictive validity	The extent to which a text-based measure predicts relevant and expected outcomes.	Text-based measures of compassion and openness should predict subsequent immigration rates in a country.

Note. This is not an exhaustive list of validities, just a list of examples showcasing how scores extracted from text should be validated against an external ground truth.

Beyond English Texts

Since language has downstream effects on supposedly non-linguistic cognitive domains (e.g., memory, social cognition, decision-making), English-centric NLP studies of historical processes could tremendously mislead researchers (see Blasi et al., 2022). This limitation inhibits applications of NLP methods in a truly inclusive psychological science. Some projects have created time-tagged corpora in multiple non-English languages, but many text-based studies assume English as the "default language," and English speakers as the "default human." English, however, is only one of the approximately 6,900 languages spoken or signed in the world today

and linguistic research has uncovered substantial diversity. Some computational linguists have also voiced major concerns about this problem in NLP (Bender, 2019). Recognizing and addressing the problem of English-centricity in text analysis (which is homologous to the WEIRD people problem in traditional psychology) is critical for psychologists using NLP techniques because failing to do so may ignore substantial linguistic diversity in the world (Atari & Dehghani, 2022). Non-English written languages that have had an overwhelming historical significance as carriers of culture over centuries (e.g., classical Chinese, Sanskrit, Persian, Greek, and Latin) can be of particular interest to historical psychology.

Notably, structured historical data may be more likely to be available and/or accessible in developed countries that have longer histories of modern institutions or in countries that were not colonized. Even within nations or time periods, language from certain social classes tend to be most represented in written documents (e.g., high-status men who were likely literate). As we mentioned earlier, culture is history in the present, so if researchers of the future mine our current social-media data for historical insight, their findings will primarily describe WEIRD people with smartphones and reliable internet access. This issue does not diminish the value of learning from these rich textual sources; rather, noting such issues will help us further enrich our analyses of cultural change.

Conclusions

Traditionally, cultural and cross-cultural psychology have primarily focused on cross-societal differences in behaviors, norms, values, and traits, ignoring cross-temporal differences within a population. The roots of cross-cultural variation, however, often lie in historical processes (Henrich, 2020; Muthukrishna et al., 2021). Metaphorically, cross-societal examinations are like an art gallery in which multiple static photographs are exhibited, while

historical psychology is like a movie, dynamically connecting those snapshots to provide context-rich insights about why, when, and how things got to where they are now. Since we do not have direct access to “dead minds” who lived in the past, we can rely on the textual records they left behind, ranging from personal notes and poetry to novels and religious inscriptions. Specific historical psychological hypotheses can come from the cumulative framework of cultural evolutionary theory, offering novel ideas about how our psychology has changed throughout historical time. Taking history seriously is a critical part of moving beyond the WEIRD people problem and making psychology a genuinely universal and inclusive psychological science.

Recommended Reading

Atari, M. & Dehghani, M. (2022). Language analysis in moral psychology. In M. Dehghani & R. Boyd (Eds.), *The Atlas of Language Analysis in Psychology* (pp. 207-228). Guilford Press.

- A review of different text-analysis methods in psychology, focusing on moral text analysis, cultural relevance of text analysis, going beyond WEIRD languages, and increasing language and researcher diversity in the field.

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. Penguin UK.

- In this book, the author explores how institutions and psychology jointly influence each other over time and how a series of Catholic Church edicts on marriage undermined the foundations of kin-based societies, leading to the emergence of WEIRD psychology in Western societies.

Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology*, *72*, 717-749.

- A proposal for psychology to become a historical science, the authors review studies that may be classified as historical psychology, introduce sources of historical data, explain the crucial role of cultural-evolutionary theory, and outline how psychologists can add historical depth and nuance to their research.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, *366*(6466), eaau5141.

- An empirical piece in which the authors studied the historical origins of WEIRD psychology, showing how the edicts by the medieval Roman Catholic Church, such as the prohibition on cousin marriage, weakened kinship ties and led to a more impersonal

social structure. In this historical psychology paper, the authors draw on insights from multiple disciplines, combining historical data on the duration of exposure to the medieval Western church, ethnographic data on the prevalence of cousin marriage and polygamy, and present-day cross-cultural data on various behavioral tendencies.

Varnum, M. E., & Grossmann, I. (2021). The psychology of cultural change: Introduction to the special issue. *American Psychologist, 76*(6), 833-837.

- This is the introductory article to a Special Issue at the journal *American Psychologist*, putting together cutting-edge research and theory to address the "what," "why," and "how" of cultural change, laying out the authors' and editors' hopes to encourage more psychologists to consider cultural change in their work.

References

Atari, M. & Dehghani, M. (2022). Language analysis in moral psychology. In M. Dehghani & R. Boyd (Eds.), *The Atlas of Language Analysis in Psychology* (pp. 207-228). Guilford Press.

Atari, M., Reimer, N. K., Graham, J., Hoover, J., Kennedy, B., Davani, A. M., … Dehghani, M. (2021). Pathogens are linked to human moral systems across time and space. *Current Research in Ecological and Social Psychology, 3,* 100060.

Bender, E. (2019). The #BenderRule: On naming the languages we study and why it matters. *The Gradient.* Retrieved from https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/

Blasi, D. E., Henrich, J., Adamou, E., Kemerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Science.* https://doi.org/10.1016/j.tics.2022.09.015

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science, 32*(2), 218-240.

Choi, E., Farb, N., Pogrebtsova, E., Gruman, J., & Grossmann, I. (2021). What do people mean when they talk about mindfulness?. *Clinical Psychology Review, 89,* 102085.

Choi, V. K., Shrestha, S., Pan, X., & Gelfand, M. J. (2022). When danger strikes: A linguistic tool for tracking America's collective response to threats. *Proceedings of the National Academy of Sciences, 119*(4), e2113891119.

Cole, M. (1990). Cultural psychology: A once and future discipline? In R. A. Dienstbier & J. Berman (Eds.), *Nebraska Symposium on Motivation: Vol. 37. Current theory and research in motivation* (pp. 279– 335). Lincoln: University of Nebraska Press.

Davani, A. M., Díaz, M., & Prabhakaran, V. (2022). Dealing with disagreements: Looking beyond the majority vote in subjective annotations. *Transactions of the Association for Computational Linguistics, 10*, 92-110.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*(1), 134-149.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), E3635-E3644.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology, 26*(2), 309-320.

Gray, R. D., Bryant, D., & Greenhill, S. J. (2010). On the shape and fabric of human history. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1559), 3923-3933.

Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science, 24*, 1722–1731.

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

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. Penguin UK.

Henrich, J., Bauer, M., Cassar, A., Chytilová, J., & Purzycki, B. G. (2019). War increases religiosity. *Nature Human Behaviour, 3*(2), 129-135.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences, 33*(2-3), 61-83.

Ihara, Y., & Feldman, M. W. (2004). Cultural niche construction and the evolution of small family size. *Theoretical Population Biology, 65*(1), 105-111.

Jackson, J. C., Watts, J., List, J. M., Puryear, C., Drabble, R., & Lindquist, K. A. (2022). From text to thought: How analyzing language can advance psychological science. *Perspectives on Psychological Science, 17*(3), 805-826.

Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2022). Text analysis for psychology: Methods, principles, and practices. In M. Dehghani & R. Boyd (Eds.), *The Atlas of Language Analysis in Psychology*. Guilford Press.

Markus, H. R., & Hamedani, M. G. (2020). People are culturally shaped shapers: The psychological science of culture and culture change. In S. Kitayama & D. Cohen (Eds.), *Handbook of Cultural Psychology* (2nd ed., pp. 11–52). Guilford Press.

Martins, M. D. J. D., & Baumard, N. (2020). The rise of prosociality in fiction preceded democratic revolutions in Early Modern Europe. *Proceedings of the National Academy of Sciences, 117*(46), 28684-28691.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour, 3*(3), 221-229.

Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology, 72*, 717-749.

Nisbett, R. E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently… and Why*. New York: The Free Press

Pinker, S. (1997). *How the mind works*. Penguin.

Rentfrow, P. J., & Jokela, M. (2016). Geographical psychology: The spatial organization of psychological phenomena. *Current Directions in Psychological Science, 25*(6), 393-398.

Scheffer, M., van de Leemput, I., Weinans, E., & Bollen, J. (2021). The rise and fall of rationality in language. *Proceedings of the National Academy of Sciences, 118*(51), e2107848118.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science, 366*(6466), eaau5141.

Slingerland, E. (2014). Toward a second wave of consilience in the cognitive scientific study of religion. *Journal of Cognitive Historiography, 1*(1), 121-130.

Slingerland, E., Atkinson, Q. D., Ember, C. R., Sheehan, O., Muthukrishna, M., Bulbulia, J., & Gray, R. D. (2020). Coding culture: Challenges and recommendations for comparative cultural databases. *Evolutionary Human Sciences, 2,* E29.

Snefjella, B., Généreux, M., & Kuperman, V. (2019). Historical evolution of concrete and abstract language revisited. *Behavior Research Methods, 51*, 1693–1705.

van Loon, A., Giorgi, S., Willer, R., & Eichstaedt, J. (2022). Negative Associations in Word Embeddings Predict Anti-black Bias across Regions–but Only via Name Frequency. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 1419-1424).

Vardy, T., & Atkinson, Q. D. (2019). Property damage and exposure to other people in distress differentially predict prosocial behavior after a natural disaster. *Psychological Science, 30*(4), 563-575.

Varnum, M. E., & Grossmann, I. (2017). Cultural change: The how and the why. *Perspectives on Psychological Science, 12*(6), 956-972.

Vygotsky, L. S., & Luria, A. R. (1993). *Studies on the history of behavior: Ape, primitive, and child.* Mahwah, NJ: Erlbaum. (Original work published 1930)

Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural evolution of genetic heritability. *Behavioral and Brain Sciences*, 1-147.

Winkler, M. (2022). *Do disasters affect the tightness of social norms*. Working paper. Retrieved from https://static1.squarespace.com/static/6012d580037eab74e6a1e7dd/t/6038022b45dcad582f8473be/1614283345510/Winkler_jmp.pdf