HISTORICAL TIME IN THE AGE OF BIG DATA
Cultural Psychology, Historical Change, and the Google Books Ngram Viewer

Michael Pettit
York University

Launched in 2010, the Google Books Ngram Viewer offers a novel means of tracing cultural change over time. This digital tool offers exciting possibilities for cultural psychology by rendering questions about variation across historical time more quantitative. Psychologists have begun to use the viewer to bolster theories about a historical shift in the United States from a more collectivist to individualist form of selfhood and society. I raise 4 methodological cautions about the Ngram Viewer’s use among psychologists: (a) the extent to which print culture can be taken to represent culture as a whole, (b) the difference between viewing the past in terms of trends versus events, (c) assumptions about the stability of a word’s meaning over time, and (d) inconsistencies in the scales and ranges used to measure change over time. The aim is to foster discussion about the standards of evidence needed for incorporating historical big data into empirical research.

Keywords: historical psychology, digital history, Google, cultural psychology, big data

The Google Books Ngram Viewer (http://books.google.com/ngrams) quickly captured the attention of social scientists, humanists, and journalists after its 2010 launch. Initially dismissed in corridor talk as a superficial toy, this digital tool has begun to seep into our methodological toolbox. The prospects are exciting. It offers a “big data” perspective on cultural trends over the course of hundreds of years. It promises to deliver a means of analyzing temporal variations in cultural psychology alongside the currently flourishing ways of assessing geographical differences. In short, there are genuine hopes for reviving a historical psychology (Staeuble, 1991) with contemporary social scientific rigor. Any such integration of history and psychology is still very much in its infancy. In September 2015, I was able to identify 11 psychological studies indexed in PsycINFO that relied on the Ngram Viewer to make empirical claims about cultural psychology. Nevertheless, certain tendencies are already apparent in this literature. I raise four methodological cautions about the Ngram Viewer’s use among psychologists when it comes to making arguments about change over historical time. My aim is not to criticize individual scholars in an underdeveloped field, but rather to help articulate a future set of best practices for evaluating claims about historical psychology.

The Ngram Viewer is not an isolated development, but is part of a larger transformation of how we study and understand the human in an age of “big data.” Sociologists Mike Michael and Deborah Lupton define big data as “massive and continually generated digital datasets that are produced via interactions with online technologies” (Michael & Lupton, 2016, p. 104). In the 21st century, “knowing capitalism” has eclipsed the social scientist’s ability to acquire information about human conduct (Savage & Burrows, 2007). One response to this transformation is for traditional disciplines to embrace computational technologies for data collection and analysis, becoming digital humanities in the process (for overviews relevant to the history of psychology, see Green, in press; Pettit, Serykh, & Green, 2015). In what follows, some of the issues I raise are general ones relating to a digital approach to historical topics. However, my primary aim is to...
offer an analysis of how psychologists have actually used the Google Books Ngram Viewer to make arguments about historical changes to the self. This is a nascent area of research with much potential, but also a number of pitfalls.

**Google Ngrams and the Promise of Historical Psychology**

The Ngram Viewer is based on the massive amount of textual material digitized by Google Books in collaboration with university libraries. This process involved scanning each page of millions of physical books and having the text digitized with optical character recognition (OCR) software. According to the scientific paper announcing the tool, Google had scanned over 15 million books. The Ngram Viewer (2010) corpus consisted of a subset of this massive archive: 5 million books selected for the quality of the OCR and the bibliographical metadata. The authors estimated that this represents approximately 4% of all books ever published (Michel et al., 2011). Google released an updated corpus in October 2012 with more titles, improved OCR, and corrected metadata (Orwant, 2012). An ngram is a single unit (or a word) and the viewer allows users to trace the frequency of different ngrams (phrases up to five words) across the span of specified years. Although the corpus includes some books dating back to the 1500s, its extent expands dramatically beginning in the 19th century, mirroring the rise of mass print culture. Similarly, there are texts available in numerous languages. However, English predominates. When conducting a search, the user can select from a numbers of languages, and for the English language, they can distinguish among English, American English, British English, and English Fiction.

The Ngram Viewer offers novel ways for psychologists to incorporate “historical time” into their theories. Historical time is one way of escaping the tired dichotomy between nature and nurture, by attending to changes in psychological dispositions and expression that take place on the midrange time frame of cultural change as opposed to the more familiar deep time of evolution or developmental time of the human life span (Pettit & Davidson, 2014; Pettit & Hegarty, 2014). Just as cultural psychologists have done much to document how perception, self, cognition, emotion, and motivation vary synchronously across space (Henrich, Heine, & Norenzayan, 2010; Markus & Kitayama, 1991), historical psychology examines diachronic variation. It can facilitate a move away from the often criticized trope of the static distinction between a holistic East and atomistic West (Adams & Markus, 2004; Rozin, 2003). At least since Wundt, leading psychologists have voiced an interest in studying psychological phenomena not only culturally, but in a historical manner that extends their laboratory-based research (Cahan & White, 1992). However, research methods textbooks invariably present the very passage of time as one of the greatest threats to the validity of psychological studies. The traditional challenge facing a temporally sensitive “eventful psychology” has been the rigorous collection and analysis of empirical evidence about the mental past. In this respect, the corpus of Google Books represents a tremendous archive of human cognitions, affects, and behaviors that is accessible and measurable in unprecedented ways.

However, the Ngram Viewer’s adoption is occurring against the backdrop of psychologists’ renewed concerns about how poor methods for data collection and analysis can seriously compromise the scientific enterprise (Eich, 2014; Ledgerwood, 2014). Despite their seductive appeal, big data analyses are far from immune from questions about replication, reliability, and validity (Lazer, Kennedy, King, & Vespignani, 2014). If psychologists want to incorporate historical time into their analysis of culture, then a discussion about proper methods is necessary. It is undeniable that historical inquiry will never achieve the level of control (e.g., the creation of randomly assigned treatment groups) that has become the norm in experimental psychology. Nevertheless, historical psychology can be informed by the comparative methods championed by political scientists (Skocpol & Somers, 1980) and longstanding interpretive practices among historians for evaluating the provenance and audience for historical texts (Jordanova, 2000).

In sum, the use and interpretation of Google Books Ngram Viewer data speak to the relationship of psychology and history in the age of big data. Psychology and history have been traveling down separate paths since the 1970s when it comes to these disciplines’ troubled identity as
social sciences. Jill Morawski (2011) has discussed the “non-debate” within psychology over qualitative and interpretivist methods. In contrast, historians have been reluctant to embrace quantitative methods favoring instead the tentative curation of unique archival documents as their primary way of knowing (Lamont, 2009, pp. 79–87). Sadly, the most discussed “manifesto” for digital methods offers a very idiosyncratic and frankly unreliable account of recent historiography (Cohen & Mandler, 2015). In the 1970s, there was a brief interest in chiometrics, for example, the turn to economics, demography, and quantitative sociology to illuminate patterns of historical change and the experience of the ordinary persons. However, incredulity about the accuracy of historical data sets, especially when it came to the politically vexed issues of the “efficiency” of slavery as a labor system (Gutman, 1975), constrained much of the enthusiasm. Poststructuralist concerns about the historical stability of categories of identity and the universality of the experiences of the lives lived under those signs led to further skepticism in the 1990s (Scott, 1991). However, this “cultural turn” left historians ill-equipped to deal with questions of structure and causality (Sewell, 2005). The digital humanities promise ways of incorporating a historical sensibility with the concerns of other social sciences. However, in our enthusiasm for new techniques, we should not forget the difficulties in acquiring reliable and valid quantitative data for the past or those interpretivist insights into the historicity of meaning and identity.

Print Culture ≠ Culture

Thus far, the most common use of the Ngram Viewer among psychologists has been to bolster theories about a shift from a collectivist to an individualist mind-set (Greenfield, 2013; Kesebir & Kesebir, 2012; Twenge, Campbell, & Gentile, 2013)—the historical change from gemeinschaft (community) to gesellschaft (society). The key assumption underlying this use is that print culture offers a representative sample of culture as a whole. This is a deeply problematic assumption. It presumes a certain level of literacy, linguistic homogeneity within nation-states, and the widespread consumption of the printed word. Each of these varies considerably across time and space. They are themselves historical phenomena deeply intertwined with the change these psychologists are trying to document. For example, every article uses the American English category to transparently represent the culture of the United States. This assumption about linguistic homogeneity within nation-states does not hold for cosmopolitan New York City or Chicago in 1900 (Barrett, 1992; Gutman, 1973), or even for the French countryside for much of the 19th century (Weber, 1976).

This assumption confuses the psychology of a society as a whole with a particular form of a sometimes high, sometimes middletorn culture. For example, the corpus includes very few examples of the cheap but largely ephemeral dime novels and pulp fiction that were widely popular with American working-class readers (Egnal, 2013). It also necessarily excludes the visual culture, oral traditions, and symbolic rituals that have been at the heart historical and anthropological interpretations of culture for decades. It assumes that cultures are undifferentiated and totalizing wholes whose self-understanding is driven by the thought leaders who make it into print. The tool provides no ways of disaggregating distinct subcultures. In short, it risks replicating a version of the past that 50 years of social history and cultural studies have debunked (Hall, 1981). Instead, we need to recognize that the Google Books corpus does not represent culture as an undifferentiated whole, but foregrounds the culture of particular subgroups.

Big data analysis often appears to offer a census of the population, but one is invariably working with a sample that still requires considerable interpretation. This situation is made worse by the fact that the researcher is often not fully aware of how sampling occurred because data collection and the corpus’s curation are proprietary knowledge (Boyd & Crawford, 2012). In the case of the Ngram Viewer, Michel and his team (Michel et al., 2011) mentioned that periodicals were excluded from the corpus, leading many scholars to assume that scientific and medical journals were not part of the sample. This is incorrect. Eitan Adam Pechenick, Christopher M. Danforth, and Peter Sheridan Dodds have demonstrated through a comprehensive statistical analysis that the corpus does not represent a random, unbiased sample of all publications (Pechenick, Danforth, & Dodds,
Instead, scientific publications are increasingly overrepresented compared with other genres in the corpus’s coverage of the 20th century. This is an unsurprising finding considering that the digitization of university libraries first generated the collection.

There are lags between historical experience and its representation in print. Green (2013) found that three words associated with the discredited psychological theory of mesmerism (mesmerism, mesmeric, mesmerist) are more common in the English Fiction of the 1990s than any time during 1800s. This finding undoubtedly reveals something about contemporary culture (such as a nostalgia for an imagined past), but it does not represent everyday behavior or norms in a straightforward manner. This is not intended as an attack on digital methods, but rather a call for us to be more pluralistic in the digital databases we use (and construct!) to represent the past and to be more modest about the historical claims being made. For example, Jean Twenge, the most prolific contributor of ngram studies, uses the tool as one aspect of a larger research program deploying a variety of social science methodologies to examine the rise of narcissistic traits in the millennial generational cohort. However, only one psychological study used other linguistic sources (e.g., dictionaries) in making their claims to complement research with the Ngram Viewer (Oishi, Graham, Kesebir, & Galinha, 2013). Acknowledging such limits when it comes to Google’s corpus will facilitate greater precision about the assumed population under scrutiny, as exemplified by recent work on trends in the American novel (Egnal, 2013).

**Secular Trends Versus Eventful History**

The Ngram Viewer is a kind of time series analysis, a long-standing tool in a number of disciplines (Klein, 1997). The viewer’s default setting is a threefold smoothing of the curve. This has the effect of making cultural change look like a biological or financial phenomenon with now-familiar secular trends. In contrast, when one sets the smoothing to zero, one sees spikes often associated with the publication of particular texts in specific years. This makes visible the historical events that often puncture and subtly shape cultural structures (Sewell, 2005). In other words, the ngram analysis could be viewed as a kind of quasi-experiment: an interrupted time series with specified historical events serving as the “treatments” leading to cultural change (Campbell, 1969). I would urge scholars to use both views and to consider what each perspective reveals and obscures. However, the viewer is designed to privilege looking at the past in terms of gradual cultural evolution. Thus far, only one study attempted to relate psychological states (emotional expression) to specific historical events (e.g., wars, the Great Depression, the baby boom) rather than the forces of modernization writ large (Acerbi, Lamos, Garnett, & Bentley, 2013).

**Words and Phrases**

The most common methodology among psychologists is searching for individual words that the authors see as sensitive to cultural change (Greenfield, 2013; Hamamura & Xu, 2015; Kesebir & Kesebir, 2012). For example, one team focused on pronoun use to trace the status of the women (Twenge, Campbell, & Gentile, 2012b), and to argue for a comparative declined in American English of collective pronouns (e.g., we, us) in favor of individualist ones (I, me; Twenge et al., 2013). Only three studies used phrases in addition to individual words (Mason, Kuntz, & McGill, 2015; Oishi et al., 2013; Twenge, Campbell, & Gentile, 2012a). Interestingly, although Twenge et al. (2012a) documented the same rise in individualist phrases as they did for pronouns (Twenge et al., 2013), the authors also found a greater persistence of communal phrases compared with communal words.

In cultural psychological research, I would suggest that the ngram tool is much more powerful when researching phrases. The social psychologist Peter Hegarty has shown how word order can both reflect and enact cultural values and gendered stereotypes (Hegarty, Watson, Fletcher, & McQueen, 2011). Searching for individual words is risky. A word’s meaning is as likely to change over time as its frequency. For example, I am not convinced that “God” meant precisely the same thing in 1800 versus 1900 versus 1950, even as the frequency of its use

---

1 One study used Amazon’s Mechanical Turk to crowdsource the selection of the words and phrases analyzed (Twenge, Campbell, & Gentile, 2012a).
diminished (Gibbs & Cohen, 2011). Indeed, it did not necessarily carry the same in meaning in different communities in 1900. Alongside, the Ngram Viewer’s frequency counts, the Oxford English Dictionary is a powerful tool for charting an English word’s changing connotations across time (Benjafield, 2014).

Even phrases can be ambiguous. In my own historical research (Pettit & Vigor, 2015), the phrase “menstrual calendar” can refer to both an individual woman’s biorhythm and the paper tool used to chart it. The valuable functions that Google provides to refine your research (https://books.google.com/ngrams/info) will not disaggregate these uses. Obviously, the two meanings are very closely related. The use of menstrual calendars as a technology to monitor women’s reproduction became intertwined with the subsequent scientific representation of women’s reproduction. However, the distinct uses will only be found when one clicks on the various date ranges under “Search in Google Books” and look at the particular phrase in the context of the original text.

The Scale(s) of History

Each time a user puts a request into the viewer, the scale of the y-axis changes. There is tremendous variation in the frequency with which relevant words appear. However, the audience is presented with the same-looking graphs time after time. Thus far, there has been no discussion about setting standards for comparisons across searches. Do we need a baseline for how frequently a word or phrase appears in the corpus before we start making claims about cultural change? We know that “the”—the most common word in the English language—represents about 5% of the entire corpus of words (Egnal, 2013). One researcher (Del Giudice, 2012) used the viewer to refute argument that there was a historical switch in the gendering of the colors pink and blue. He searched for “blue for boys,” “pink for girls,” “blue for girls,” and “pink for boys” and did not find the previously theorized shift. However, the key phrase “blue for boys” represents at its peak only 0.00000500% of the corpus for that year. In other words, none of the phrases he selected seem particularly representative of actual cultural usage.

There was also considerable variation in date ranges selected across studies. Two studies used the default 1800 to 2000 setting (Greenfield, 2013; Oishi et al., 2013). Del Giudice (2012) used 1880 to 1980. Kesebir and Kesebir (2012) opted for 1900 to 2000. Twenge et al. chose 1960 to 2008 in two studies (Twenge et al., 2012a, 2013), and 1900 to 2008 for a third (Twenge et al., 2012b). In the case of Twenge et al. (2013), the use of “I” was most frequent during the 19th century, and then declined between 1900 and 1960, after which its frequency increases again. Greenfield (2013) is unique in offering a fulsome explanation for focusing on a chosen time period. She uses the period from 1800 to 2000 to highlight the 1920s, the period when the United States shifted from a rural to an urban society. In a follow-up study examining cultural change in China, Zeng and Greenfield (2015) used a different date range (1970 to 2008) to reflect a similar period of modernization in that country. It is perfectly reasonable that addressing different questions will require a focus on different time periods. However, none of the other articles offered a robust justification for the particular start or end dates they selected. Readers need to be mindful of which dates a study includes and what this choice may disclose or obstruct.

Cultural History and Eventful Psychology

There has been much talk of late that historical and cultural researchers need to finally abandon their antiquated microscopes for shiny new telescopes. This rhetoric is misplaced, misrepresents how history has been practiced in the recent past, and risks needlessly impoverishing future research. At their best, the digital humanities allow one to move more easily among scales of analysis, to both zoom out and zoom in (Kaiser, 2012). There have been a number of recent examples of such mixed methods within the history of psychology itself. Christopher Green’s (2015) article on the disunity of psychology draws on his historical research on the linguistic networks his team detected in early psychology journals alongside the close reading of other primary sources. Hunter Heyck’s (2015) analysis of the postwar “age of system” builds on a digital survey of specified keywords in flagship social sciences journals between 1925 and 1975. The book also includes a
thoughtful methodological appendix for those considering similar approaches.

There are ways of drawing upon insights from the cultural history of the self to help design studies using the Ngram Viewer as a digital tool. Cultural psychologists have tended to favor comparisons between Eastern and Western cultures, but historians prefer other units of analysis. John Carson (2007) made a compelling case that the civic cultures of France and the United States led to the coproduction of different versions of “intelligence” and its administration. Figure 1 charts the use of first-person singular and plural pronouns in the French corpus (excluding the popular “on,” as this word can mean either “one” or “we”) to offer a comparison with the existing work by Greenfield (2013) and Twenge et al. (2013) on personal pronouns in American English. I expect that this analysis disproportionately represents Parisian developments, but it likely includes the words of French Canadian, Haitian, Southeast Asian, and African writers. Inspired by questions raised by an award-winning political history of the self during the French Revolution (Goldstein, 2005), the graph begins in 1700 rather than the default setting. Indeed, usage of singular pronouns spikes during the revolution (peaking in 1793) and continues through the Empire and Bourbon monarchy. After this, the plural pronoun dominates until the 1980s, long after France experienced industrialization, urbanization, and secularization. Although never a particularly common term, the more specific phrase *le moi* peaks in the 1840s at a time when Jan Goldstein argues that pedagogue Victor Cousin designed a curriculum to instruct the male bourgeoisie in the cultivation of “the self” (see Figure 2). In contrast to studies based on the American English corpus, one does not see a transition from plural to singular pronouns with the advent of modernity. It is unclear whether this reveals telling cultural differences or is an artifact of how and which books got sampled.

Given its mix of technical and popular publications, the corpus furnishes an opportunity to consider the historical place of psychology’s vocabulary within English-language discourse. Thus far, this issue has been broached using qualitative methods (Danziger, 1997) as well as the quantitative analysis of the PsycINFO and Oxford English Dictionary databases (Benjafield, 2012, 2013, 2014). One can visualize how “behavior” came to replace “conduct” by 1940 as the preeminent way of discussing human

---

**Figure 1.** The pronouns “Je+je+J’+j’+me+moi” versus “Nous+nous” in the Google Books Ngram Viewer French corpus, 1700–2000. See the online article for the color version of this figure.
activity well beyond the confines of disciplinary psychology (see Figure 3). “Opinion” and “opinions” remain a prevalent way of speaking about subjective human viewpoints (as well as a judge’s official decision), but social scientists’ interest in measuring “attitudes” of ordinary persons (Igo, 2007) contributed to this term’s increasing presence in the corpus after 1900 (see Figure 4). Over the 20th century (and especially since the publication of the DSM–III in 1980; Leys, 2000; Young, 1995), there is an increased tendency to describe a very difficult or violence experience in terms of “trauma” (see Figure 5).

As historians have argued (D’Emilio & Freedman, 1998), the United States seems to

![Figure 2. The term “le moi” in the Google Books Ngram Viewer French corpus, 1700–2000. See the online article for the color version of this figure.](image)

![Figure 3. The terms “behavior” versus “conduct” in the Google Books Ngram Viewer English language corpus, 1800–2000. See the online article for the color version of this figure.](image)
have had multiple “sexual revolutions” or at least a much longer one than popular culture typically allows. The Ngram Viewer detects a number of quantifiable changes in the words now associated with sexuality. Discussions of “sex” in all its facets become more prominent in the late 1920s, and then again and more dramatically from 1968 to 1976, in the American English corpus (see Figure 6). The widespread adoption of “gender” as a distinct category from “sex” occurs between 1983 and 1993 as second-wave feminists joined the ranks of the academy (Rutherford & Pettit, 2015). One needs to be cautious about potential gaps between discourse and practice. The works in the American English corpus relevant to sex is almost exclusively

Figure 4. The terms “opinion + opinions” versus “attitude + attitudes” in the Google Books Ngram Viewer English language corpus, 1800–2000. See the online article for the color version of this figure.

Figure 5. The term “trauma” in the Google Books Ngram Viewer English language corpus, 1880–2005. See the online article for the color version of this figure.
authored by educators, reformers, and scientists. Experts often deploy language to transform individuals and conduct into problems. The greater historical prevalence of the word “homosexuality” compared with “heterosexuality” is striking (see Figure 7), but not surprising to those versed in the history of sexuality inspired by Michel Foucault (Pettit & Hegarty, 2014).

Memory offers a curious case. “Memory” is a much older concept than organized psychology—one with a long, rich cultural life (Danziger, 2008). The term is especially prevalent in the corpus before 1700, although this result is likely because of the paucity of books from those years, which would lead to a publishing event (e.g., the appearance of a single volume...
Books that substantively mention memory before 1800 are most often either volumes dedicated to instructing the reader in the arts of mnemonics or examples of monody, a poem dedicated to the memory of specified individual. The frequency of memory compared with other terms in the corpus declines in the first half of the 20th century, only to increase again after 1960 (see Figure 9). Such usage fits well with a familiar story: the cognitive revolution that reintroduced the psychological study of mental states after decades of neglect. However, a click on the “Search in Google Books” function reveals that the increase is also due to the academic study of the “collective memory.”
of war, revolution, and mass violence. Thus far, the study of memory as individual information processing and as the shared recollection of traumatic events has largely operated in isolation from one another. However, their history is intertwined (Winter, 2012).

My methodological cautions about the Google Books Ngram Viewer are meant to open the possibility of cultural psychologists collaborating with cultural historians. Thus far, these two fields have had virtually nothing to say to one another. The silence is striking. Because technological changes have made forms of historical research seem more empirical and objective, cultural psychologists are starting to revive the notion of a historical psychology. What this field’s best practices will entail remains open. I concur with Pechenick et al. (2015) that one needs to be extremely wary about making inferences about sociocultural evolution based on analyses of this corpus. However, a consequence of their conclusion that “the corpus is in effect a library” is that this data set is much closer than first presumed to the kinds of evidence historians have studied and around which they have developed sophisticated interpretative practices.

References


