Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

Tracking the Consumption Junction

Wevers, Melvin; Gao, Jianbo; Nielbo, Kristoffer L.

document version
Publisher's PDF, also known as Version of record

Link to publication in KNAW Research Portal

Tracking the Consumption Junction: Temporal Dependencies between Articles and Advertisements in Dutch Newspapers

Melvin Wevers[1], Jianbo Gao[2], Kristoffer L. Nielbo[3]

[1]Digital Humanities Lab, KNAW Humanities Cluster, Netherlands
[2]Institute of Complexity Science and Big Data Technology, Guangxi University, China
[3]Center for Humanities Computing Aarhus, Aarhus University, Denmark

Corresponding author: Kristoffer L. Nielbo , `kln@cas.au.dk`

Abstract

Historians have regularly debated whether advertisements can be used as a viable source to study the past. Their main concern centered on the question of agency. Were advertisements a reflection of historical events and societal debates, or were ad makers instrumental in shaping society and the ways people interacted with consumer goods? Using techniques from econometrics (Granger causality test) and complexity science (Adaptive Fractal Analysis), this paper analyzes to what extent advertisements shaped or reflected society. We found evidence that indicate a fundamental difference between the dynamic behavior of word use in articles and advertisements published in a century of Dutch newspapers. Articles exhibit persistent trends that are likely to be reflective of communicative memory. Contrary to this, advertisements have a more irregular behavior characterized by short bursts and fast decay, which, in part, mirrors the dynamic through which advertisers introduced terms into public discourse. On the issue of whether advertisements shaped or reflected society, we found particular product types that seemed to be collectively driven by a causality going from advertisements to articles. Generally, we found support for a complex interaction pattern dubbed the consumption junction. Finally, we discovered noteworthy patterns in terms of causality and long-range dependencies for specific product groups. All in, this study shows how methods from econometrics and complexity science can be applied to humanities data to improve our understanding of complex cultural-historical phenomena such as the role of advertising in society.

I   INTRODUCTION

1.1   Advertisements as a Lens on the Past

Over the course of the twentieth century, branded consumer goods turned into an integral part of society Cross [2000], de Grazia [2005]. This sparked research into consumer goods, as well as research that relies on consumer goods as entry points into broader cultural phenomena. Historians, for instance, have studied how advertisements represented consumerism, gender identities, and the globalization of food cultures Lears [1994], Parkin [2007], Sivulka [2012]. For these scholars, advertisements functioned as a lens on the past. Roland Marchand argues that advertisements provide an insight into the ideals and aspirations of past realities. They show the state of technology, the social functions of products, and provide information on the society in which a product was sold. Furthermore, Marchand poses that advertisements contributed to the shaping of a "community of discourse." He claims that advertisements infused public discourse with a particular type of language Marchand [1985].

At the same time, scholars have debated to what extent adverts actually offer a meaningful depiction of the past. In their effort to sell more products, producers, and ad makers amplified or distorted certain social and cultural aspects to make products more appealing to consumers Brandt [2004]. Erving Goffman points out that ads shaped consumer's lived experience by prescribed certain conceptions of identity Goffman [1985]. Put differently, ad makers used adverts to *shape* society by tweaking the desires of consumers Fox [1997]. This form of social engineering was rooted in psychological and sociological theories of the day. Especially the tobacco industry was very actively involved in developing new ways of advertising that could entice people to pick up cigarettes Brandt [2004].

These theoretical debates raise the question whether we can actually use advertisements to study the past, or are we merely studying a version of the past constructed by ad makers? Theories on the relationship between advertisements and society can be summarized by three positions. The first position contends that advertisements reflected the desires and aspirations of consumers. The second argues that advertisements merely represented the interest of advertisers and the companies that produced the commodities. This approach attributes more agency to the advertisers. The third approach acknowledges that there has existed a more complex "consumption junction", in which producers, distributors, consumers, and advertisers collectively negotiated the meaning and success of a consumer product Cowan [1997].

This article examines the validity of these three positions in a specific historical context (i.e. twentieth-century Netherlands) and sets out to answer to what extent advertisements reflected or shaped society. We studied this interplay between advertisers and society by analyzing advertisements and articles in newspapers. Newspapers are a well-read rich historical source that contains "conscious representations of conditions and events" as well as "unconscious reflection[s] of the tastes, the interests, the desires, and the spirit of its day [Smail, 2008, p.48]." As such, newspapers function as a proxy for public discourse Schudson [1982], Marshall [1995]. Moreover, the availability of digitized newspapers enables researchers to use computation to explore the archive, locate particular instances of language use, or extract specific linguistic patterns.

This study combines techniques from econometrics and complexity science to examine the dynamics of word use in articles and advertisements. We set out to answer the following three questions. First, did advertisements shape or reflect newspaper discourse? Second, did word use in ads differ in dynamics from articles? And finally, were these characteristics more pronounced for particular product groups? Answering these questions will further our understanding of the function of advertising in society.

1.2 Beyond Counting Word Frequencies

With the proliferation of large databases that hold temporally-dispersed text content, time series plots of word frequencies have become a valuable source of exploration and validation of cultural trends Gao et al. [2012], Michel et al. [2011]. Google has popularized this approach through their web-based n-gram viewer of time-dependent relative word frequencies derived from its digitized book collection.[1] The popularity of the Google Ngram viewer has sparked other digital archives to also develop n-gram viewers based on their collection.[2]

---

[1] Google Books Ngram Viewer: https://books.google.com/ngrams

[2] Examples of ngram viewers: The British Library: https://www.webarchive.org.uk/ukwa/ngram/, Danish Royal Library: http://labs.statsbiblioteket.dk/smurf/, and National Library of the Netherlands http://kbkranten.politicalmashup.nl

N-gram viewers can help researchers to determine when words appeared and how they evolved. Plots of word frequencies, however, only offer an overview of particular trends in discourse. To gain more insight into the trends in and between advertisements and articles, we applied two techniques from econometrics and complexity science to the (relative) frequencies of words. First, we applied the Granger causality test to determine whether trends in word use in ads impacted trends in articles or vice versa. Second, we analyzed the persistence of words using fractal analysis to identify whether ads differed from articles in terms of dynamics related to word use. Put differently, we determined whether advertising discourse was distinct in its behavior from discourse in articles. Also, we set out to identify specific words that *stuck* with people? The later analysis could point towards the existence of what has been called communicative memory.

1.3 Communicative Memory

Communicative memory is one of the many concepts in the field of memory studies. It has been linked to heterogeneous concept cultural memory Donald [2001], Assmann [2011]. Jan Assman has described cultural memory as "a collective concept for all knowledge that directs behavior and experience in the interactive framework of a society and one that obtains through generations in repeated societal practice and initiation [Assmann and Czaplicka, 1995, p. 125]." One of these repeated societal practices is language use, which also shapes our collective understanding of a shared culture. According to Assman, cultural memory is formed over large periods of time whereas communicative memory represents memories shaped over shorter time spans (80-100 years). In a way, communicative memory can be viewed as the short-term memory of a society Assmann [2008], Assmann and Czaplicka [1995].

In this paper, communicative memory is approximated by time series of word frequencies newspapers for a period of a century. Memory in a time series is modeled as the presence of self-similarity, or more precisely persistent correlation, between the values of these features at various time steps. By modeling time series of word use in newspapers, we try to capture expressions of communicative memory in public discourse. Merely glancing at visualizations produced by n-gram viewers might show sudden peaks or slow decays in word use. In this paper, however, we quantify such behaviors and show the extent to which word use exhibits particular memory functions.

II METHODS

2.1 Data

The National Library of the Netherlands (KB) has digitized thousands of Dutch historical newspapers using optical character recognition (OCR) software.[3] This software turns scans of physical pages into machine-readable data. Unfortunately, the text extracted from the digital scans is often flawed due to imperfections in the original material or limitations of the recognition software. These material blemishes cause the software to not recognize and transcribe every word correctly, which has resulted in conjoined words, complete gibberish, or words in which certain characters were replaced. The age and quality of the original material are important determinants of the software's ability to correctly recognize the text; hence, older newspapers contain many more errors than more recent papers. For this reason, we focused on twentieth-century newspapers. Also, the KB does not provide suitable metrics on the quality of the OCR'ed text

---

[3] These newspapers can be accessed through Delpher: https://www.delpher.nl

Traub et al. [2015]. The study, therefore, assumes that OCR errors are uniformly distributed over the period.

For analysis, we relied on two subsets of the digitized newspaper data. The first subset consisted of the entire set of advertisements ($n_1 = 18,564,411$) in the KB's digitized newspaper archive. The second set held newspaper articles ($n_2 = 11,465,220$) from two national newspapers: *De Tijd* (1890-1974) and *De Telegraaf* 1893-1989. During digitization, the OCR software separated articles from advertisements and stored the document type in the metadata, allowing us to select these two types of documents. Because advertisements made up a smaller portion of the newspapers, we selected the entire set of advertisements to make it more comparable to the corpus of articles in terms of size. Also, we narrowed our focus to two national newspapers because these are more likely to represent wider public discourse than regional newspapers.

We calculated keyword frequencies, more specifically, normalized relative daily term frequency per document for these two subsets. We explicitly looked at 265 words (singular and plural forms collapsed) that denoted consumer products. Based on exploratory data analysis using an n-gram viewer, we selected words that appeared in both advertisements and articles throughout greater periods of time with considerable frequency. Brand names were excluded for two reasons. First, brand names often appeared as part of logos in advertisements, making it more difficult to convert these images to machine-readable text. Second, the techniques used in this paper necessitate the existence of the same words over longer periods of time. More often than not there existed multiple brand names for the same products, which were also not used over longer periods of time.

2.2 Causal Dependencies

Several techniques can be used to compare lagged values of time series X with values of a second time series Y to model variation in their correlation coefficient as a function of temporal displacement. The most widely used technique is cross-correlation, which is simply used to detect the variation in the correlation between two time series as a function of lag. The Granger causality test goes beyond mere correlation Granger [1969] and tests for the existence of causal-like dependencies between temporally disjunctive time series of, for instance, words from two sources. The test, which originated in econometrics, is based on the assumption that causality is more than temporal disjunction, it involves directionality between time series. The relation tested by the Granger causality test is often characterized as predictive causality and represented as X *Granger cause* Y to distinguish it from more direct causality Sugihara et al. [2012]. At its core, Granger causality, which is related to correlation, expresses if values of time series X contain information that is uniquely predictive of subsequent values in time series Y.

For our study, we used the Granger causality test as follows. To identify a *shaping* relation, we test if variation in a specific word frequency for newspaper discourse (Y) at time t is predicted by variation in the frequency for the same word in advertisement discourse (X) at earlier time steps $t - 1 \ldots t - k$. We test for X *Granger cause* Y, by comparing the performance of the nested 'newspaper discourse only' model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_k y_{t-k} + \epsilon \tag{1}$$

4

with the full 'newspaper and advertisement discourses' model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_k y_{t-k} + \alpha_1 x_{t-1} + \ldots + \alpha_m x_{t-m} + \epsilon \tag{2}$$

to identify which one does the better job at explaining y_t based on the residuals. The zero-model for the hypothesis then is $H_0 : \alpha_i = 0$ for each i of the element $[1, m]$ with the alternative hypothesis being $H_1 : \alpha_i \neq 0$ for at least one i of the element $[1, m]$. We applied the test bi-directionally such that a shaping relation finds support if we can confirm that '$X\ Granger\ cause\ Y$' and reject that '$Y\ Granger\ cause\ X$' in case of a reflecting relationship (the inverse of shaping). Finally, if both '$X\ Granger\ cause\ Y$' and '$Y\ Granger\ cause\ X$' find support this is viewed as support for a more complex relationship between the two time series.

2.3 Long-range Dependencies

In addition to the Granger Causality test, we used fractal analysis to identify if words exhibited long-range dependencies.[4] Long-range dependency indicates a rate of decay between two points with increasing time intervals that is slower than exponential decay. Analysis of time-dependent change in complex systems—systems composed of many interacting elements—is an important application of fractal analysis.

Some random processes in complex systems are self-affine, that is, fluctuation patterns at shorter time scales are (statistically) similar to fluctuations at longer time scales. In the case of reading, for instance, fluctuations in reading speed are self-affine across multiple time scales, because both reading fluency and comprehension are affected by elements at short time scales (e.g., words and sentences) and longer times scales (e.g, paragraphs and chapters) OBrien et al. [2014]. Such fractal behavior is found in a range of culturally relevant processes related to psychology Chater and Brown [1999], economy Marchant [2008], sociology Gao et al. [2017], health Eke et al. [2002], language Gao et al. [2012] and music Voss and Clarke [1975]. We argue, therefore, that Fractal analysis has great potential for the study historical trends in cultural expressions. This is particularly the case when we are dealing with 'big data' consisting of large sets of mostly unknown parameters Gao et al. [2012].

We are interested in a particular kind of fractal processes called $1/f^{2H+1}$ processes, in which H refers to the Hurst exponent. The Hurst exponent quantifies the degree of long-range dependencies in a time series, such that when $0 < H < 0.5$, the time series is an anti-persistent process (i.e., a jump up is followed by a jump down, or vice versa, in the increment process), when $H = 0.5$, the time series only has short-range dependencies, and when $0.5 < H < 1$, the time series is characterized by long-range dependencies (i.e. a jump up is followed by another jump up, or vice versa, in the increment process), see figure 1. It is possible the $H > 1$ indicates a non-stationary process. In this study, persistence represents whether a word 'stuck' with people and it is in that manner analogous to communicative memory.

---

[4] Long-range dependency is also called persistent behavior or long-memory in time series. The terms will be used interchangeably.
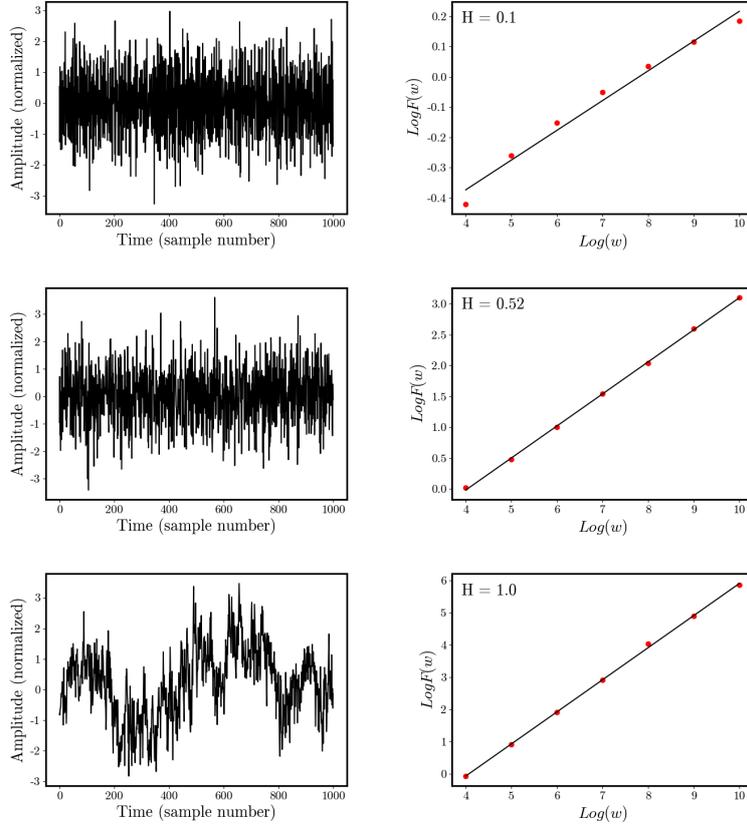
Figure 1: Left: Time series that exhibit anti-persistent (top), short-range (middle), and long-range (bottom) dependencies. Right: Estimation of the Hurst exponent for the matching time series in the left column.

2.3.1 Adaptive Fractal Analysis

Adaptive fractal analysis (AFA) is a relatively new technique for determining the Hurst exponent of a time series Gao et al. [2011], Riley et al. [2012]. AFA improves the popular detrended fluctuation analysis Peng et al. [1994] by identifying a global smoothed trend that can automatically deal with arbitrary, strong nonlinear trends Gao et al. [2011]. The technique is based on a nonlinear adaptive multi-scale decomposition algorithm Gao et al. [2011].

After constructing a random walk process from the time series, the initial step of AFA involves partitioning the time series into overlapping segments of length $w = 2n + 1$, in which neighboring segments overlap by $n + 1$ points. In each segment, the time series is fitted with the best polynomial of order M, obtained using standard least-squares regression. The fitted polynomials in overlapped regions are then combined to yield a single global smoothed trend. Denoting the fitted polynomials for the $i - th$ and $(i + 1) - th$ segments by $y^i(l_1)$ and $y^{(i+1)}(l_2)$, respectively, where $l_1, l_2 = 1, \cdots, 2n + 1$, we define the fitting for the overlapped region as

$$y^{(c)}(l) = w_1 y^{(i)}(l + n) + w_2 y^{(i+1)}(l), \ \ l = 1, 2, \cdots, n + 1, \tag{3}$$

6

where $w_1 = \left(1 - \frac{l-1}{n}\right)$ and $w_2 = \frac{l-1}{n}$ can be written as $(1 - d_j/n)$ for $j = 1, 2$, and where d_j denotes the distances between the point and the centers of $y^{(i)}$ and $y^{(i+1)}$, respectively. Note that the weights decrease linearly with the distance between the point and the center of the segment. Such a weighting is used to ensure symmetry and to effectively eliminate any jumps or discontinuities around the boundaries of neighboring segments. As a result, the global trend is smooth at the non-boundary points and it has the right and left derivatives at the boundary Riley et al. [2012].

The global trend can be used to maximally suppress the effect of complex nonlinear trends on the scaling analysis. The parameters of each local fit are determined by maximizing the goodness of fit in each segment. The different polynomials in the overlapped part of each segment are combined such that the global fit will be the smoothest fit of the overall time series. Even if the local fits are linear, $M = 1$, the global trend signal will still be nonlinear. AFA then can be described accordingly: for an arbitrary window size w, we determine, for the random walk process $u(i)$, a global trend $v(i), i = 1, 2, \cdots, N$, where N is the length of the walk. The residual of the fit, $u(i) - v(i)$, characterizes fluctuations around the global trend, and its variance yields the Hurst exponent H according to the following scaling equation:

$$F(w) = \left[\frac{1}{N} \sum_{i=1}^{N} (u(i) - v(i))^2\right]^{1/2} \sim w^H. \tag{4}$$

By computing the global fits, the residual, and the variance between original random walk process and the fitted trend for each window size w, we can plot $\log_2 F(w)$ as a function of $\log_2 w$. The presence of fractal scaling amounts to a linear relation in the plot, with the slope of the relation providing an estimate of H, see figure 1.

2.4 Design

To determine whether advertisements reflected or shaped public discourse, we first applied Granger causality tests to each of the 265 keywords, comparing time series from newspaper and advertisement discourse. We hypothesize the existence of the following three causal-like patterns:

1. Advertisements shaped newspaper articles as expressed by Granger causality directed exclusively from advertisements to articles;
2. Advertisements reflected newspaper articles as expressed by Granger causality directed exclusively from articles to advertisements;
3. A complex, possibly externally-driven, causal pattern as evidenced by cases where Granger causality goes from articles to advertisements *and* vice versa.

For the second step of the analysis, we used AFA to model the persistence for each keyword in both discourses. This enabled us to identify possible dynamic properties of the two types of discourse as a whole and possibly the dynamic properties of particular words. Similar to causal-like patterns, the Hurst exponent has three possible patterns of persistence: anti-persistent processes, short-term correlation processes, and persistent processes. Each keyword's behavior can thus be described by one of nine possible combinations of causality (Granger causality test) and persistence (AFA). Insights into these dynamic properties alongside the causal patterns can help to increase our understanding of the relationship between advertisements and articles, and by extension, between advertisements and society.
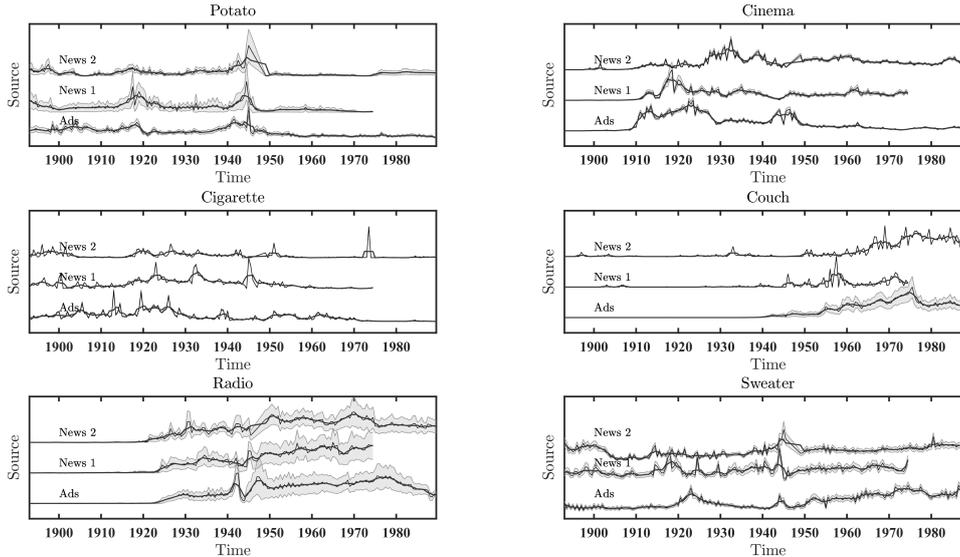
Figure 2: Six keyword frequencies plotted at bi-annual intervals for the two newspapers (News 1: *De Tijd*; News 2: *De Telegraaf*) and the advertisements (Ads). The line is smoothed using a simple moving average filter (window size of five years) and gray bands represent confidence intervals at 95%

2.5 Data Analysis

Statistical tests were conducted with an α level of .005 Benjamin et al. [2017]. Pearson's correlation coefficient R was used to measure the non-lagged association strength between the time series. We converted Pearson's R using Fisher's Z-transformation to normally distributed z-values to permit averaging. Before applying the Granger causality test for comparison of discourses, lag-1 differencing was used to obtain a stationary keyword time series.

For the analysis of the Hurst exponent for each keyword time series per discourse, we used a simple linear regression and compared this with the constant model. This allowed us to test differences in long-range dependencies between the two different discourses. The Shapiro-Wilk W test confirmed that the distribution of the Hurst exponent did not deviate significantly from normality Shapiro and Wilk [1965].[5]

III RESULTS

3.1 Directionality

On average, the variance in correlation for each keyword in all the time series was similar between advertisement and articles. The mean correlation coefficient, $\bar{R}$, between advertisements and the newspapers *De Tijd* and *De Telegraaf* was $\bar{R} = .25$ and $\bar{R} = .27$ respectively. Sixty-two percent (62%) of these correlations were statistically reliable. The within-discourse correlation, that is, the correlation between *De Tijd* and *De Telegraaf*, was considerably stronger, $\bar{R} = .42$. Seventy-three percent (73%) of these correlations were significant, suggesting that word use over time in articles between these two newspapers was more similar than between the articles and advertisements.

---

[5] While some keyword time series did show indications of multifractal structure (i.e. local fluctuations with either small or large variation), this information was discarded from the final analysis for the purpose of simplification.

Analysis showed that there was no overarching causal pattern, but rather multiple causal patterns that were keyword-dependent. Twenty percent (20%) of the product terms show evidence of a shaping causality, in which discursive trends in advertising discourse uniquely predict those in articles. For 17% of the product terms, we found the causal pattern in which advertisements reflect articles. Almost half of the product terms (49%) belong to the complex externally-driven category, that is, both advertisements and articles seem to be primarily impacted by extraneous events that perturb the reflecting-shaping dynamic between the time series. Finally, 14% of the terms show no indication of predictive causality.

Almost half of the product terms exhibit complex behavior that points towards an external source driving changes in discourse in advertisements and newspapers. These could include economic developments or possibly the invention of new products or technology. In a way, this indicates that the relationship between advertisements and articles was one of negotiation with external developments. Noteworthy keywords in this category were related to produce ('apple', 'cauliflower', 'lettuce'), energy ('stove', 'cokes', 'furnace', 'gasoline') and audiovisual technology ('tape recorder', 'gramophone', 'radio', 'television').

There are slightly more keywords for which advertisements were shaping articles than ads reflecting articles. In case of shaping, we detected that behavior in word use in advertisements was related to behavior in articles. Word that exhibited this behavior referred quite generally to fashion and clothing ('men's clothing', 'sweater', 'fur', 'wool', 'flannel', 'jeans', 'heels'), interior design ('living room', 'dining room', 'bedroom') and movies ('cinema', 'film'). The keywords that exhibited reflecting behavior were more difficult to categorize into particular categories. They ranged from words such as 'teapot', 'dictionary', to 'cheese'. Generally, the words with reflecting behavior seemed to be more specific than the words in the shaping category. For instance, the keywords 'chair cushions' and 'winter coat' are specific types of cushions and coats.

3.2 Detection of Communicative Memory

H for both articles ($t_{1058} = 32.8, p < .00001$) and advertisements ($t_{538} = 38.5, p < .00001$) was significantly higher when compared to a no-memory baseline (H: $M = 0.5$, $SD = 0.18$). In terms of persistence, discourse in articles and advertisements was different from processes that only showed short-range correlations. AFA found an average difference (ΔH) between articles of 0.21. The dynamics related to word use in articles thus differed from advertisements. Word use in articles was more likely to be a persistent process (H: $M = 0.89$, $SD = 0.19$) than it was in advertisements (H: $M = 1.1$, $SD = 0.17$). Notice that advertisements display non-stationary dynamics with $H > 1$. To test whether the difference between ads and articles was significant, we ran linear regression to predict H as a function of advertisement and articles (advertisement as baseline). Compared to the constant model, a statistically significant regression model was found ($\chi^2_1 = 149.1, p < .00001$) showing that H was, indeed, reliably lower for articles than for advertisements (articles: β = -0.18 , SE ±0.01, $F_{1,793} = 163.9 \ p < .00001$).

On the whole, word use in articles more clearly expressed persistent trends, while word use in advertisements tended to be more irregular, displaying bursts of high activity followed by little or no activity. This indicates that articles more closely express behavior that could be interpreted as communicative memory, while ads seem more haphazard and perhaps catalytic to the construction of communicative memory. This finding resonates with Marchand's claim that advertisements contributed to the "shaping of a 'community of discourse,' an integrative common language shared by an otherwise diverse audience [Marchand, 1985, p.XX]."

9

Persistence in art.	Persistence in ads & art.	No persistence
living room	cigarettes	cinema
dining room	fur	film
bedroom	wool	tape recorder
chair	flannel	gramophone
couch	jeans	radio
cupboard	heels	television
seat	apple	
lamp	cauliflower	
	lettuce	
	cokes	
	furnace	
	gasoline	

Table 1: Example of keywords grouped on type of persistent trend (persistence in articles only, persistence in articles and advertisements, and no persistence

Our findings indicate three distinct types of persistent trends, see table 1. Persistence in only articles, persistence in advertisements and articles, and a lack of persistence in either source. Words that only exhibited persistence in articles included products related to interior design ('living room', 'couch', 'lamp', 'bedroom'). This suggests that discourse about these products was part of a shared language but not clearly part of an advertising discourse. Conversely, words that showed persistence in both discourse included those related to cigarettes but not cigars and tobacco ('cigarettes'), fashion ('fur', 'jeans'), energy ('cokes', 'furnace', 'gasoline') and produce ('apple', 'cauliflower', 'lettuce'). This suggests that advertising discourse for these products was much more persistent and relied on an established frame of reference. Interestingly, cigarettes and fashion are often presented as typical examples of strongly branded products Blaszczyk [2011], Hill [2002], White et al. [2012], Brandt [2004]. Advertisements for produce, on the other hand, might exhibit persistent processes due to its highly seasonal and reoccurring nature. Keywords that showed no persistence were related to technology ('cinema', 'tape recorder', 'gramophone', 'radio').

IV   DISCUSSION

Using AFA, we found a significant difference in persistent behavior between word use in advertisements and articles. The latter exhibited long-term dependencies whereas advertisements displayed more non-stationary and irregular behavior. In general, advertisements introduced terms, but many of these terms did not persist and their decay was rapid. For articles, on the other hand, keywords that denoted products showed more persistent behavior and were either mentioned recurrently in a self-reinforcing manner or decayed much slower than advertisements. We speculate that this reflects an overarching media dynamic in which ads introduced keywords and articles represented how these products became part of public discourse. However, this dynamic does not hold for all products as evidenced by table 1 and, at least partially, by keywords that exhibit a reflecting causal pattern.

The time series of keywords between the two newspapers were more clearly correlated than between the newspapers and advertisements. This shows that word use in newspapers more closely followed each other than word use between advertisements and newspapers. Along with the found different in H between articles and advertisements, the difference in correlation adds evidence to the hypothesis that the dynamics of discourse in ads are different from articles.

This suggests that advertisements are not merely a lens on the past, but more clearly a distorted mirror that is shaped to a certain degree by advertisers and its own dynamics.

In terms of directionality, we did not find one dominant pattern. For 20% percent of the keywords, advertisements reflected articles, and for 17% of the keywords, advertisements shaped articles. But for almost half of the keywords, there was a more complex causal relationship, indicative of external forces. This lends support to Cowan's argument for a complex interaction pattern in which advertisers, distributors, producers, and consumers negotiated the meaning of a product Cowan [1997].

4.1   Product Groups

The causal direction and type of persistence seems to be, to some extent, related to the type of product. We were not able to identify specific categories of keywords in the *reflecting* causal category. However, the complex relationship and *shaping* category offered interesting groupings of words. The groupings made on the basis of the existence of memory and causal directionality leads to the following four points of discussion.

First, products with a shaping dynamic *and* long-term dependencies in articles might point towards products that are not constantly advertised — expressed as the lack of persistence in ads — but that nevertheless are part of the cultural life of Dutch consumers throughout the twentieth century, such as bikes, pets, interiors, and clothes. The shaping dynamic reveals that ad makers might have pushed the popularity of these products, which can be described as lifestyle products. One could argue that advertisers might have been able to affect the longevity of these products, effectively installing them within communicative memory.

Second, one of the most noteworthy behaviors is associated with the cigarette. This product exhibits persistence in advertisements *and* in articles, and it shows a shaping causal behavior. This suggests that in advertising discourse and newspaper discourse, cigarettes were a recurring topic that built upon earlier discourse. Moreover, advertisers seemed to be able to shape newspaper discourse on cigarettes. This finding is in line with scholars that view advertisements for cigarettes as the prime example of social engineering Brandt [2004]. Our study finds that, at least for the Dutch context, cigarette advertisements were a steadily successful form of advertising. The unique behavior of cigarettes was underlined by the fact that related products such as tobacco and cigars behaved dissimilarly. Tobacco and cigars exhibit no persistence and are driven by a complex causal relationship, underscoring different advertising dynamic than found for cigarettes.

Third, some products revealed persistence in both advertisements and articles without displaying a uniform causal relationship. These products include produce, energy sources, and computer systems. One interpretation might be that produce was of prolonged importance (indicated by the existence of long-range dependencies), yet its importance was not driven by advertisers but by an external source. In the case of produce, this external source could be seasonal or economic shifts. The other two product groups (energy sources and computer systems) were also quite instrumental in society, albeit not for prolonged periods during the twentieth century. The keywords associated with energy were most dominant in the first half of the century, whereas, the words associated with computers only appeared in the latter quarter of the century. Nonetheless, they both still exhibited persistence in newspaper articles.

Finally, keywords associated with technological innovations showed two distinct types of behavior. First, keywords such as 'cinema', 'tape recorder' and 'television' did not exhibit any

persistence, which could have been the results of the constant innovation and disruptions in the field of audiovisual technology. Another explanation could be the use of different keywords to refer to similar technologies. Further research is needed to explore this behavior related to technology. Second, we found a distinction in the causal relationship between types of technology. 'Cinema' and 'film' showed a clear causal relationship between ads and articles. The causal relationship might have resulted from the fact that advertisements played an important role in pushing these innovations to a wider audience. Keywords associated with household technology ('radio' and 'television'), on the other hand, displayed the complex type of causality. These technological products might be more closely related to particular economic, seasonal, or innovative cycles. Again, further research is needed to untangle these dynamics.

V CONCLUSION

Through two data experiments, we have found evidence of a fundamental difference between the dynamic behavior of word use related to consumer products in articles and advertisements published in newspapers. Articles — taken as a proxy for public discourse — exhibit persistent trends that are likely to be reflective of communicative memory. Contrary to this, advertisements have a more irregular behavior characterized by short bursts and fast decay, which, in part, mirrors the dynamic through which advertisers introduced terms into public discourse. On the issue of whether advertisements shaped or reflected society, we found particular product types that seemed to be collectively driven by a causality going from advertisements to articles. Generally, we found support for a complex interaction pattern dubbed the consumption junction. Finally, we discovered noteworthy patterns in terms of causality and long-range dependencies for specific product groups.

This study shows how methods from fields of econometrics and complexity science can be applied to improve our understanding of complex cultural-historical phenomena. Further research that employs more extensive keyword lists that also includes brand names and cross-cultural comparisons will make it possible to propose a more detailed and general account of the mechanics that underpin this consumption junction.

5.1 Acknowledgements

References

Jan Assmann. Communicative and Cultural Memory. In Astrid Erll, Ansgar Nnning, and Sara B. Young, editors, *Cultural Memory Studies: An International and Interdisciplinary Handbook*, pages 109–119. Walter de Gruyter, New York, 2008. ISBN 978-3-11-018860-8.

Jan Assmann. *Cultural Memory and Early Civilization: Writing, Remembrance, and Political Imagination*. Cambridge University Press, Cambridge ; New York, December 2011. ISBN 978-0-521-76381-3.

Jan Assmann and John Czaplicka. Collective Memory and Cultural Identity. *New German Critique*, 65:125, 1995. ISSN 0094033X. doi: 10.2307/488538. URL http://www.jstor.org/stable/488538?origin=crossref.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, Eric-Jan Wagenmakers, Richard Berk, Kenneth A. Bollen, Bjrn Brembs, Lawrence Brown, Colin Camerer, and others. Redefine statistical significance. *Nature Human Behaviour*, 2017. URL https://dukespace.lib.duke.edu/dspace/handle/10161/15438.

Regina Lee Blaszczyk, editor. *Producing Fashion: Commerce, Culture, and Consumers*. University of Pennsylvania Press, Philadelphia, October 2011. ISBN 0-8122-0605-3.

Allan Brandt. Engineering Consumer Confidence in the Twentieth Century. In Xun Zhou and Sander Gilman, editors, *Smoke: A Global History of Smoking*. Reaktion Books, London, 2004. ISBN 1-86189-200-4.

Nick Chater and Gordon DA Brown. Scale-invariance as a unifying psychological principle. *Cognition*, 69(3): B17–B24, 1999.

Ruth Schwartz Cowan. A Proposal for Research Strategies in the Sociology of Technology. In Wiebe E. Bijker, Thomas Hughes, and Trevor Pinch, editors, *The Social Construction of Technological Systems*, pages 261–280. MIT Press, Cambridge, 1997. ISBN 0-262-52137-7.

Gary Cross. *An All-Consuming Century: Why Commercialism Won in ¡odern America*. Columbia University Press, New York, 2000. ISBN 0-231-11312-9 978-0-231-11312-0 978-0-231-11313-7 0-231-11313-7.

Victoria de Grazia. *Irresistible Empire: America's Advance Through Twentieth-Century Europe*. Belknap Press of Harvard University Press, Cambridge, kindle edition, 2005. ISBN 0-674-01672-6.

Merlin Donald. *A Mind So Rare: The Evolution of Human Consciousness*. W. W. Norton & Co., New York, August 2001. ISBN 978-0-393-04950-3.

A. Eke, P. Herman, L. Kocsis, and L. R. Kozak. Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23(1):R1, 2002. URL http://stacks.iop.org/0967-3334/23/i=1/a=201.

Stephen R. Fox. *The Mirror Makers: A History of American Advertising and its Creators*. University of Illinois Press, Urbana, 1997. ISBN 0-252-06659-6 978-0-252-06659-7.

Jianbo Gao, Jing Hu, and Wen-wen Tung. Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering. *PLoS ONE*, 6(9):e24331, September 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0024331. URL http://dx.plos.org/10.1371/journal.pone.0024331.

Jianbo Gao, Jing Hu, Xiang Mao, and Matja Perc. Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. *Journal of The Royal Society Interface*, 9(73):1956–1964, 2012.

Jianbo Gao, Peng Fang, and Feiyan Liu. Empirical scaling law connecting persistence and severity of global terrorism. *Physica A: Statistical Mechanics and its Applications*, 482:74–86, September 2017. ISSN 03784371. doi: 10.1016/j.physa.2017.04.032. URL http://linkinghub.elsevier.com/retrieve/pii/S0378437117303357.

Erving Goffman. *Gender Advertisements*. Macmillan, London, 1985.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969. URL http://www.jstor.org/stable/1912791.

Daniel Delis Hill. *Advertising to the American Woman, 1900-1999*. Ohio State University Press, Columbus, 2002. ISBN 978-0-8142-0890-8.

T. J. Jackson Lears. *Fables of Abundance: A Cultural History of Advertising in America*. BasicBooks, New York, 1994. ISBN 0-465-09076-1 0-465-09075-3.

Roland Marchand. *Advertising the American Dream: Making Way for ¡odernity, 1920-1940*. University of California Press, Berkeley, 1985.

Thierry Marchant. Scale invariance and similar invariance conditions for bankruptcy problems. *Social Choice and Welfare*, 31(4):693–707, December 2008. ISSN 0176-1714, 1432-217X. doi: 10.1007/s00355-008-0298-8. URL http://link.springer.com/10.1007/s00355-008-0298-8.

Margaret Marshall. *Contesting Cultural Rhetorics: Public Discourse and Education, 1890-1900*. University of Michigan Press, Ann Arbor, 1995. ISBN 0-472-10536-1.

J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176–182, January 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1199644. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1199644.

Beth A. OBrien, Sebastian Wallot, Anna Haussmann, and Heidi Kloos. Using Complexity Metrics to Assess Silent Reading Fluency: A Cross-Sectional Study Comparing Oral and Silent Reading. *Scientific Studies of Reading*, 18(4):235–254, July 2014. ISSN 1088-8438, 1532-799X. doi: 10.1080/10888438.2013.862248. URL http://www.tandfonline.com/doi/abs/10.1080/10888438.2013.862248.

Katherine Parkin. *Food Is Love: Advertising and Gender Roles in Modern America*. University of Pennsylvania Press, Philadelphia, January 2007. ISBN 978-0-8122-1992-0.

C.-K. Peng, Sergey V. Buldyrev, Shlomo Havlin, Michael Simons, H. Eugene Stanley, and Ary L. Goldberger. Mosaic organization of DNA nucleotides. *Physical review e*, 49(2):1685, 1994.

Michael A. Riley, Scott Bonnette, Nikita Kuznetsov, Sebastian Wallot, and Jianbo Gao. A tutorial introduction to adaptive fractal analysis. *Frontiers in Physiology*, 3, 2012. ISSN 1664-042X. doi: 10.3389/fphys.2012.

00371. URL http://journal.frontiersin.org/article/10.3389/fphys.2012.00371/abstract.

Michael Schudson. *The Power of News*. Harvard University Press, Cambridge, 1982. ISBN 978-0-674-69586-3.

Samuel Sanford Shapiro and Martin B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. URL http://www.jstor.org/stable/10.2307/2333709.

Juliann Sivulka. *Soap, Sex, and Cigarettes: A Cultural History of American Advertising*. Wadsworth, Boston, 2nd edition, 2012. ISBN 0-534-51593-2 978-0-534-51593-5.

Daniel Lord Smail. *On Deep History and the Brain*. University of California Press, Berkeley, 2008. ISBN 978-0-520-25289-9.

G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. Detecting Causality in Complex Ecosystems. *Science*, 338(6106):496–500, October 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1227079. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1227079.

Myriam C. Traub, Jacco van Ossenbruggen, and Lynda Hardman. Impact analysis of OCR quality on research tasks in digital archives. In *International Conference on Theory and Practice of Digital Libraries*, pages 252–263. Springer, 2015.

Richard F. Voss and John Clarke. 1/fnoise in music and speech. *Nature*, 258:317, November 1975. URL http://dx.doi.org/10.1038/258317a0.

Cameron White, John L. Oliffe, and Joan L. Bottorff. From the Physician to the Marlboro Man Masculinity, Health, and Cigarette Advertising in America, 19461964. *Men and Masculinities*, 15(5):526–547, December 2012. ISSN 1097-184X, 1552-6828. doi: 10.1177/1097184X12461917. URL http://jmm.sagepub.com/content/15/5/526.