

Understanding mediated practices: Combining ethnographic methods with blog data to develop insights

Jonathan Bean

Bucknell University

Zeynep Arsel

Concordia University

While theories of practice have been influential in the social sciences, these frameworks have seen limited application in ethnographic and applied inquiry, perhaps because few methods for carrying out practice theoretical research have been elaborated. We address this opportunity and provide an account of a multi-method inquiry on domestic practice. First, we explain methods for integrating data from blogs with ethnographic methods and how this data can be used to develop theory. Second, we share our experience as interdisciplinary researchers using ethnographic and quantitative data to connect work at the boundaries of social practice theory and theories of consumption. Finally, we share our insights on why industry should aim to better understand existing and emergent consumer practices.

INTRODUCTION

While theories of practice have been influential in the social sciences, much applied and empirical consumer research tends to investigate the choices and actions of discrete individuals or groups of consumers instead of analyzing the routines, engagements and performances that constitute social life. In contrast to the common usage of the term, such as in the phrase “best practices,” theoreticians of social practice contend that structure, agency, and the dynamic relationship between individuals and the market should be the starting point for research. Therefore, from this point of view, consumers and consumption should be investigated through the study of practice: study not cell phone users, but cellphoning (Shove, Watson, Hand, and Ingram 2007). A common heuristic for the analysis of practice is a tripartite scheme, variously described as objects, meanings, and doings (Magaudda 2011, Arsel and Bean 2013) stuff, images, and skills (Scott, Bakker, and Quist 2012), or equipment, images, and competencies (Shove and Pantzar 2005). Despite the promise of the practice theoretical approach in applied contexts, these frameworks have seen limited application in ethnographic and applied inquiry, in part because few methods for carrying out practice theoretical research have been elaborated. Furthermore, although linking up ethnographic studies with broader patterns of practice could help to better illuminate cultural patterns, few ethnographic and applied researchers combine ethnographic methods, such as long interviews, with the visual and textual data generated and used on sites such as Pinterest, Facebook, and popular blogs. We contend online media can provide data on the patterning and distribution of existing and emergent practices that would

otherwise be difficult to ascertain through traditional ethnographic practice. Our paper addresses these two opportunities by outlining a method for dealing with the large amount of text and image data found on a popular blog.

CONTEXT

The example used in this paper is Apartment Therapy, which started in 2004 as a blog and has since become a media brand focusing on domestic consumption. Whereas Martha Stewart invokes picket fence perfectionism (Golec 2006), the aesthetic of Apartment Therapy is soft modernism, a blend of the elitist form-follows-function ethos of high modernism that incorporated the popular preference for restrained use of color and the pursuit of comfort and warmth (Gebhard 1995). For example, Le Corbusier and Mies van der Rohe represent high modernism; Crate and Barrel and IKEA use soft modernism to sell home goods to middle-class consumers. To these consumers, soft modernism serves an important economic function by winnowing a consumer's choices to a set that is not only more manageable, but also more likely to meet with broader acceptance — and thus garner higher value when a home is resold (Rosenberg 2011). As a central conduit for the communication of soft modernism, Apartment Therapy appeals to the young, relatively affluent consumer actively seeking advice on all aspects of domestic practice. Started as a home design blog, it quickly established sites focused on cooking, parenting, home technology, and green living, emerging as a powerful media force with greater reach than Martha Stewart Living, Sunset, and other well-read shelter magazines.

The challenge of observing soft modernism at work

The method we discuss emerged from a logical necessity to incorporate content from the Apartment Therapy blog site into an ethnographic study on the readers of the blog. The rich narrative and visual imagery of blogs well complements ethnographic analysis because blogs are public and spontaneous representations of everyday practice (Arsel and Zhao 2013).

These narratives are encoded in mass mediated representations and are shared stories that are typically understood as representing and providing the meaning component of a practice. They can be complex and even contradictory, such as that surrounding the Hummer SUV brand in the US (Luedicke, Thompson, and Giesler, 2010), or simple, such as a shared understanding that an Apple iPod makes an ideal gift. As the iPod example makes clear, however, cultural narratives also can influence actions — in this case, the giving of iPods. This, in turn, can create shifts in the alignments of objects, meanings, and doings that constitute a practice (Magaudda 2011). Therefore, it is essential to understand the range and trajectory of broader cultural narratives and to incorporate this understanding into ethnographic analysis.

One way of thinking of mediated cultural representations is to see them as maps of the circuit of a practice, where the boundaries and limitations of the circuit are drawn, held and tested. One might expect a blog like Apartment Therapy to employ a complex system of editorial guidelines and approval procedures, but a surprise early in the research was the lack

of centralized oversight over the discourse. At the time of our research, however, the bulk of Apartment Therapy's content came from a collection of freelance bloggers who were paid per post. Rather than edit and approve individual posts, Apartment Therapy employed a tryout process where interested bloggers would submit a series of sample posts. These posts would run on the blog and editors would choose which freelancer to hire permanently. To be hired, a freelancer had to exhibit not only an affinity to the linguistic style of the blog, but also a familiarity with the aesthetic sense of soft modernism. Thus, we contend that the posts on Apartment Therapy and similar blogs can be read as the material evidence and expression of the embodiment of an organic, self-referential, and continuous narrative.

As one might imagine, the quantity of information on blogs, while not approaching that of big data, can overwhelm typical qualitative methods of analysis. To address this problem, we provide below practical guidance on how to collect, store, organize and analyze large and potentially messy data sets. First, we describe a method used to automate the extraction and formatting of a database of nearly 2 gigabytes (about 55,000 blog posts) of text and image data. Second, we walk the audience through the use of database software to archive and organize the textual and visual content of blogs. Third, we discuss how to use natural language processing software to generate and analyze a corpus of textual blog content. Fourth, we show how this process can bolster qualitative analysis and help to identify illustrative sample posts from the collected data.

Extracting data from blogs

The first step is to create an offline archive of blog content. Before starting any kind of analysis, it is essential to clean the data to avoid clutter and to increase efficiency. This can be done in one of two ways: either manually — by a human manipulating a computer and saving each blog post to a separate electronic file — or it can be automated to some extent with a program that automatically download all web content by pointing the program at the blog's archive pages, which list all past posts by month or by category. Automating the process of creating an offline archive, however, may take some ingenuity, especially because server-side blog software has changed, and not all blogs have easily accessed archives. With the advent of so-called “endless” scrolling and image-driven microblogging formats such as Tumblr, there may be some unavoidable manual labor to create a list of links to all past posts. That said, you may be able to find software such as TumblrRipper, which can be used to create offline archives of image content. Note that it is a violation of the terms of service of many commercial blogs to download large amounts of content, so your project may require special permission of the blog's owner.

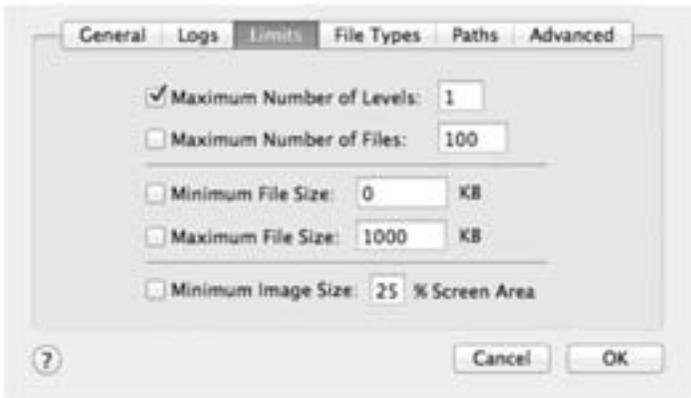


FIGURE 1. LIMITS. The *Limits* tab on SiteSucker's settings panel, shown here, allows the user to specify how many levels are downloaded. Setting the maximum number of levels to 1 means that the program will download the web page to which it is pointed and all content linked to on that page, then stop.

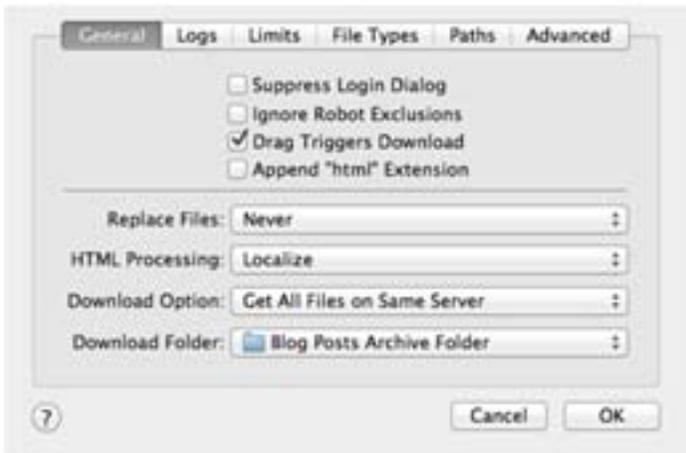


FIGURE 2. OPTIONS. The *General* Tab, also in SiteSucker's preferences, allows you to specify if the program should download files only on the same server to which it is pointed, or if it should download files regardless of location on the web. Localize, the default setting for HTML Processing, changes code in the files you download so images can be viewed offline and pages link to the offline version rather than back to the original server.

If you cannot find a way to use software that will make a useful archive of blog content, the simplest way to automate the download of a large number of web links is to create a web

page using a free online service such as Google Sites or Wordpress.com. First, assemble the list of URLs that you wish to download, such as links to each blog index page or individual blog post. Once you have this list, use Excel's text concatenation function to transform the links into valid HTML code by using a formula similar to the one below. If the links are in column A, the formula in cell B1 would read:

=CONCATENATE("", A1, "")

Once you fill row B with this formula, you will have a list of web links in HTML format that can be copied and pasted into the web page you made on Google Sites or Wordpress.com. From there, use a program such as SiteSucker for Mac to download the contents of each web link. Be careful when setting the options so that you get the correct amount of data. In particular pay attention to settings that allow you to specify the number of levels or layers of link hierarchy that you want to retrieve. If you are studying a site with relatively few links to outside content, it might make sense to download one extra level of content. For example, this could allow you to archive other blog posts and websites referenced in links in posts in the focal blog. On the other hand, if your site is large or complex, downloading links that lead away from the original domain may result in an unmanageable mountain of data; for your first round of analysis we suggest that you download only data (text and images) that appear in posts on your target domain. After some time, the automated download software will produce a set of separate HTML files. Make a backup copy of these HTML files before proceeding with the next steps.

Because nearly all web pages have some repeated content, such as a text or graphic header, navigation, search functions, and advertising, in most cases the downloaded files will benefit from some further processing to remove repetitive content. (If not, you may find yourself wondering why "search," "next," or "author" are the most commonly used words in your dataset.) Using a computer tool called grep, which can be likened to a much more powerful version of the familiar search and replace function found in a word processing program, is the fastest way to do this across many text files. If you do not possess skills using grep or HTML, or wish to gain them, we recommend you bring on outside help. A savvy computer science student can make quick work of the next steps.

Using grep, compose and test a search expression that will strip out the non-desired repetitive content (headers, advertisements, navigation, and so on) from your web page. Most blogging platforms generate well-organized and commented code, often divided with the "<div>" html tag, so it is relatively easy to identify which sections you need to strip out and devise a grep expression accordingly. For example, the header section of most HTML files is used for things such as advertising tracking and will appear on every HTML page you have downloaded. It is likely this section is not needed for your analysis. The precise grep string you would use will vary with the site, but essentially you want to tell the computer to look for a string of text that starts and ends the section. Often HTML code will be commented to denote the sections of the page. These comments, which start with "<!--", can be useful as anchors for a grep expression.

Once you have tested the grep expression on several files, you can apply it to all of the HTML files you have collected using the command line or with a batch search-and-replace

function in an application such as TextWrangler for Mac. If you find that an error in your grep code or an irregularity in the HTML files you have collected has caused a problem, you can revise as necessary and run the script again on your backup. Once you have a clean set of files you can proceed with manual analysis or put the files in a database, which can help ease the process of retrieval and comparison.

For database software, we recommend DevonThink Pro Office, which runs only on Mac OS, for several reasons.

- **Built-in RSS feed support.** RSS stands for really simple syndication... which can help automate the addition of new blog posts over time
- **An artificial intelligence engine.** Select one database item, and DevonThink can find database items that are linguistically similar to it. A central finding in our research — and the centerpiece of a diagram explaining our theory — is, in emic terms, the “landing strip.” This term refers to an area near the entry to your home where you stash keys, mail, shoes, and outerwear to keep them from “contaminating” the rest of your living space. We identified blog posts on the topic of a landing strip and used this function to find other posts that used similar language, which helped us to extend and test our theory and also allowed us to find the most illustrative examples for our readers without reading the thousands of posts containing the term “landing strip.”
- **Sophisticated searches.** In addition to Boolean searches (searching for one word AND another word OR some other word but NOT this word), DevonThink can find records containing words that are near each other. For example, to find evidence to back up a preliminary finding that Apartment Therapy linked the state of cleanliness to the feeling of calm, we looked for instances of the word “calm” near the word “clean.” These searches can be saved and as new information comes into the database they will automatically be included in the search results.
- **Redundancy.** DevonThink has a built-in backup function.
- **Export tools.** Records in DevonThink can be easily exported to a variety of formats, including text, for further processing in specialized text analysis software.
- **AppleScript support.** AppleScript is a macro instruction language that allows repetitive tasks to be automated. We used an AppleScript and grep to change the creation date of each record in our database to match the date and time it was originally published on the AT blog. This allowed us to easily sort chronologically and divide the posts into groups according to date.

A primary disadvantage of DevonThink Pro Office is that its power and sophistication taxes a computer’s resources. If you plan to work with a database as large or larger than ours, we recommend using the fastest computer you can and making sure it has as much random access memory installed as possible. In particular this will speed complex searches. A second disadvantage is that DevonThink Pro Office is not designed for collaboration. While it is possible to save the database file using a web sharing service such as Dropbox, only one user can open and work with the database file at a time. If it is opened by two or more users at

once, it is highly likely that the file will become corrupted and you will lose data. To solve this problem, we developed a system of “checking out” the file for access via a quick email notification, and as a secondary measure we used the built-in label function of Mac OS X to turn the file red to show that it was not available for use. When one user was done working with the file, we changed its Finder label to green to show it was available for the use of others. While other database or qualitative analysis software can be used to the same effect, but we found that the combination of DevonThink’s ease of use and power makes it particularly well suited to researchers who are beginning to work with large datasets.

SYNTHESIZE. GO BACK AND BACKWARDS. REPEAT.

Social science knowledge is created through an iterative process that requires the investigator to go back and forth between data and provisional understandings of the phenomenon. Multi-method inquiries necessitate another layering of analysis, to merge and synthesize various types of data at different levels of specificity. This back-and-forth shifting of focus is essential because it broadens the analysis from emic to etic and synthesizes multiple interpretations, including the researcher’s own observations, the subjects’ probed narratives, and the subjects’ performed narratives (such as blog content). It also broadens the temporal and spatial spectrum of inquiry from the resource restricted boundaries of ethnographic inquiry.

While there is no universal rule to where to start, our subjective opinion is to begin by analyzing the ethnographic data first. This, we feel, holds closer to the spirit of ethnographic inquiry by allowing the researcher to begin to frame analysis based on the words and observed actions of the research participants. This can help prevent the researcher from relying too much on the somewhat decontextualized analysis a machine can make. While powerful in extracting patterns, machine generated analysis has a significant blind spot in terms of losing some of the nuances and context (Maxwell 2013). We advocate an approach where insights from interviews, content analysis, coding, or other ethnographic methods comes first, before the use of automated tools.

Once some provisional understanding is achieved, the researcher can perform analysis on the entire body of data to test and develop emergent theories. No single software is fit for all uses; in our case we chose the Stanford Part-of-Speech Tagger (Toutanova et al. 2003) because its capabilities mapped our provisional findings. We observed that the categories of objects, meanings, and doings could be roughly mapped onto parts of speech: objects were typically represented by nouns, meanings by adjectives, and doings by verbs. Using the Stanford Part-of-Speech Tagger allowed us to efficiently find the most commonly used nouns, adjectives, and verbs in our corpus. Note that the software does not do the analysis for you; as a check measure, we independently coded the results as objects, meanings, and doings, and discussed inconsistencies until they were mutually resolved.

After you have invested the time in getting your data into usable shape, it is worth investigating other analytic tools. For example, if your analysis aims to show the change in time of personal relationships, network analysis may be useful. To gain a sense of difference in how people are using words over time, software tools such as WORDij can quickly reveal shifts. If you are interested in the emotional content of a corpus of text, the Linguistic

Inquiry and Word Count software can be of use. Riopelle (2013) discusses how ethnographers can use these specific tools in an applied context. Humphreys (2010) utilizes a similar multi method approach, supplementing qualitative coding with computer-assisted content analysis to trace the changes in discourses on gambling in American newspapers.

CONCLUSION

As Warde (2005) has famously argued, consumption is only one moment in practice. Taken together, ethnography and social practice theory have the potential to illuminate both applied and academic studies of consumption. For the consultant, applying the social practice framework can show how integral practice is to the purchase and use of goods and services (Korkmann 2006), therefore offering a theoretical framework that is both substantial enough to contain the complexity of the world, but also comprehensible enough to be used by members of corporate teams regardless of previous exposure to theory. For academic audiences, applying social practice theory is one way to address the challenge of connecting potentially related research in different contexts — in effect, to show how practices intersect and draw from one another (Shove et al 2012). For example, we showed how a key resource for 21st century consumers recruited into the Apartment Therapy taste regime was the practice of soft modernism, a practice that had roots in postwar cultural changes. For those visiting the Apartment Therapy site, soft modernism provided not only a source of objects — teak coffee tables, sleek lamps, Danish chairs — but also a source of meaning. Together, objects and meanings are critical resources that become inseparably integrated with the doings (the third element of practice) of soft modernism in a way that would become closely associated with both Apartment Therapy and our participants' own putatively personal sense of style (Arsel and Bean 2012). Breaking down practice into these three constituent parts and showing how practice performances (Shove et al 2012) arise from the patterned combination of objects, meanings, and doings could lend some interchangeability to otherwise disparate analyses.

Mediated representations of practice can be an ideal site to confirm and understand the workings of practices observed by the ethnographer in the field. While a netnographic (Kozinets 2010) or critical visual (Schroeder 2006) approach can be used to understand mediated representations, our approach differs in applying the three-part practice heuristic of objects, meanings, and doings. Mapping representations of practice in these three categories can provide insights into the material workings of a practice and may provide hints on how a particular practice may change in the future.

ACKNOWLEDGEMENTS

The authors thank Maxwell Gillingham-Ryan for allowing them to use the Apartment Therapy site content for data analysis. Funding for this research project was provided by the Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC).

REFERENCES CITED

- Arsel, Zeynep and Jonathan Bean
2013 "Taste Regimes and Market-Mediated Practice." *Journal of Consumer Research* 39(5):899-917.
- Arsel, Zeynep and Xin Zhao
2013 "Blogs." In *The Routledge Companion to Digital Consumption*. Russell Belk and Rosa Llamas, eds. Pp 53-61. New York: Routledge.
- Gebhard, David
1995 "William Wurster and His California Contemporaries: The Idea of Regionalism and Soft Modernism." In *An Everyday Modernism: The Houses of William Wurster*. Marc Treib, ed. Pp. 164-183. Berkeley, CA: University of California Press.
- Golec, Michael J.
2005 "Martha Stewart Living and the Marketing of Emersonian Perfectionism." *Home Cultures* 3(1):5-20.
- Humphreys, Ashlee
2010 "Semiotic Structure and the Legitimation of Consumption Practices: The Case of Casino Gambling." *Journal of Consumer Research* 37 (3):490-510.
- Korkmann, Oskar
2006 "Customer Value Formation in Practice: A Practice-Theoretical Approach." Ph.D. diss., Swedish School of Economics and Business Administration.
- Kozinets, Robert
2010 *Netnography: Doing Ethnographic Research Online*. London: Sage.
- Luedicke, Marius, Craig Thompson, and Markus Giesler
2010 "Consumer Identity Work as Moral Protagonism: How Myth and Ideology Animate a Brand-Mediated Moral Conflict." *Journal of Consumer Research* 36 (6):1016-1032.

Magaudda, Paolo

2011 "When Materiality 'Bites Back': Digital Music Consumption Practices in the Age of Dematerialization." *Journal of Consumer Culture* 11 (1): 15-36.

Maxwell, Chad R.

2013 "Accelerated Pattern Recognition, Ethnography, and the Era of Big Data." In *Advancing Ethnography in Corporate Environments: Challenges and Emerging Opportunities*. Brigitte Jordan, ed. Pp. 175-192. Walnut Creek, CA: Left Coast Press.

Riopelle, Ken

2013 "Being There: The Power of Technology-based Methods." In *Advancing Ethnography in Corporate Environments: Challenges and Emerging Opportunities*. Brigitte Jordan, ed. Pp. 38-55. Walnut Creek, CA: Left Coast Press.

Rosenberg, Buck Clifford

2011 "Home Improvement: Domestic Taste, DIY, and the Property Market." *Home Cultures* 8 (1): 5-23.

Schroeder, Jonathan

2006 "Critical Visual Analysis." In *Handbook of Qualitative Methods in Marketing*. Russell Belk, ed. Pp. 303-321. Aldershot, UK: Edward Elgar.

Scott, Kakee, Connie Baker, and Jaco Quist

2012 "Designing Change by Living Change." *Design Studies* 33 (3): 279-297.

Shove, Elizabeth and Mika Pantzar

2005 "Consumers, Producers and Practices." *Journal of Consumer Culture* 5 (1): 43-64.

Shove, Elizabeth, Mika Pantzar, and Matthew Watson

2012 *The Dynamics of Social Practice: Everyday Life and How it Changes*. London: Sage.

Shove, Elizabeth, Matthew Watson, Martin Hand, and Jack Ingram

2007 *The Design of Everyday Life*. Oxford, UK: Berg.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer

2003 "Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network." In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*. Pp 173-80. Edmonton, Canada: Association for Computational Linguistics.

Warde, Alan

2005 "Consumption and Theories of Practice." *Journal of Consumer Culture* 5 (2): 131-53.

WEB RESOURCES

DevonThink Pro Office

<http://www.devontechnologies.com/products/devonthink/devonthink-pro-office.html>.

Stanford Log-linear Part-Of-Speech Tagger

<http://nlp.stanford.edu/software/tagger.shtml>

SiteSucker for Mac

<http://www.sitesucker.us/mac/mac.html>