

Online Research Methods, Quantitative

KRISTIN PAGE HOCEVAR and ANDREW J. FLANAGIN

University of California, Santa Barbara, USA

Communication research in the last few decades has been transformed by the wide-scale adoption and use of online research tools. Information and communication technologies have provided methodological benefits in the conduct of research, including advantages in data collection, enhanced data accuracy, and a greater capacity to effectively reach populations of interest. These and related developments largely mark improvements in the efficiency of the performance of research as existing methods increasingly migrate to online environments. As such, the use of online research tools can in many cases be viewed as an evolution of existing methods to online delivery, implementation, and data collection efforts, rather than as the creation of fundamentally new research methods.

By contrast, a host of recent technological developments has made possible essentially new means of conducting research of online populations and phenomena. Information and communication technologies have provided novel means of observing, recording, and analyzing human attitudes and behaviors that for the most part were unavailable to researchers until recently, in the forms and on the scale currently possible. These developments signal relatively new methods for research, reaching beyond mere efficiency gains in the conduct of research. As such, the use of certain online research tools can also be viewed as transformative to the research endeavor itself, as new methods are being invoked in the study of human behaviors that occur both on- and offline.

Regardless of whether online research methods are being employed to enhance the efficiency of more traditional research endeavors (e.g., the online delivery of a questionnaire, versus a paper-and-pencil version) or if they are being used in more novel ways (e.g., the natural language processing of events discussed in real time on social media across millions of users), core methodological concerns remain crucial for the conduct of sound research. For example, sampling and data integrity issues are critical concerns in the assessment of research results, irrespective of the specific methods employed.

This entry thus considers quantitative online research methods by discussing (a) developments in existing methods that improve the efficiency of research, (b) recent technological developments that have made possible essentially new means of conducting research of online populations and phenomena, and (c), in each case, the fundamental methodological concerns that remain crucial for the conduct of sound research. In addition, we conclude by considering potential research contributions to be made by communication scholars that are rooted in the use of existing and novel quantitative research methods online.

The International Encyclopedia of Communication Research Methods. Jörg Matthes (General Editor),

Christine S. Davis and Robert F. Potter (Associate Editors).

© 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc.

DOI: 10.1002/9781118901731.iecrm0174

Conducting traditional social science research online

Generally, the opportunity to conduct traditional quantitative social science research online enables more efficient procedures and enhanced data integrity for experiments and surveys, which for the most part share the same methodological concerns when conducted online as offline (Palomares & Flanagan, 2005; Reips, 2002; Reips & Krantz, 2010). For example, mechanized or automated data collection can increase data integrity over manually entered data, and data can be collected quickly and often relatively inexpensively online. Additionally, technology-mediated administration of surveys and experiments can increase design flexibility over offline administration. For example, the enhanced ability to build in skip logic, which allows the researcher to control the questions or treatment that the participant sees depending on desired factors (e.g., answers to prior questions or various experimental conditions) can be useful (Smyth & Pearson, 2011), as can the potential for integration of audio or visual material (Reips & Krantz, 2010). Although many of these features can also be built into experiments or surveys administered electronically offline, use of the Internet can increase ease of access to and data collection from populations of interest to communication scholars.

Methodological concerns and traditional quantitative methods

Among the primary methodological issues that arise when conducting social science research online are concerns about sampling and data integrity. A significant benefit of sampling populations online is that the Internet can afford access to a sample that is more representative of the general population than more frequently used offline populations such as undergraduate students (Hughes, 2012; Reips, 2002; Skitka & Sargis, 2006). Additionally, when the target population of interest in a study is Internet users, online recruitment can often be easier and more efficient than offline recruitment (Mazur, 2010; Reips, 2002). Indeed, there currently exist standing panels of potential study participants through online crowdsourced marketplaces such as Amazon.com's "Mechanical Turk," an Internet platform that enables the recruitment and payment of people to perform tasks requiring human intelligence and discrimination. The use of subjects provided through Mechanical Turk is convenient, inexpensive, and straightforward. Although the representativeness of Mechanical Turk samples has been questioned by many scholars, research suggests that respondents recruited via this tool are slightly more representative of the U.S. population than most convenience samples (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011).

Nonetheless, the use of online samples—including those drawn from Mechanical Turk—can suffer from important methodological violations. For example, online samples are often less representative than many large Internet panels or national probability samples, since online respondents tend to be younger and more liberal (Berinsky et al., 2012; Buhrmester et al., 2011; Skitka & Sargis, 2006). Additionally, a

greater risk of sampling error can be experienced with Internet studies, since there is no comprehensive list of Internet users from which to draw probability samples (Tuten, 2010). Consequently, calculating response rates can be problematic in online studies (Palomares & Flanagin, 2005), although sampling via methods such as the use of an email address list rather than other online methods (e.g., use of website ads for recruitment) can enable accurate response rate calculations (Smyth & Pearson, 2011).

Core data integrity concerns in both online and offline research include social desirability bias, self-selection, and response rates. Generally, online studies can have a decreased social desirability response bias over offline studies due to increased perceived anonymity and confidentiality (even in studies of interaction) as well as lack of direct, personal contact with researchers (Hughes, 2012; Reips & Krantz, 2010). Self-selection bias can be a problem for online studies because participants may choose to join studies after recruitment by methods such as ads on relevant websites (Scherpenzeel & Bethlehem, 2011). However, undergraduate participation for course credit, which is common in much social science research, may be even less voluntary and more problematic due to its limited representativeness (Palomares & Flanagin, 2005). Finally, response rates to Internet surveys can vary significantly, but are often low (Smyth & Pearson, 2011), although there is little evidence that lower response rates across different modes of survey administration influences the validity of findings (Skitka & Sargis, 2006). In sum, invoking traditional methods of experimental and survey data collection online can benefit researchers in a number of ways as long as these underlying methodological concerns are sufficiently addressed.

Novel research affordances of the Internet

In addition to the efficiencies afforded by the Internet to traditional research methods, novel research capabilities and interests specific to the Internet and the communities it enables have also emerged with increased Internet use. An important set of research opportunities has arisen in parallel with the growth of “Big Data,” or technology-driven data that is characterized by high data volume, rapid velocity of communication, and an incredible variety of information forms (see McAfee & Brynjolfsson, 2012). Big Data repositories naturally arise from the tremendous number of individuals using a host of websites, interactive web-based tools, and social media technologies, as people interact and contribute information in these venues. The digital and mobile communication traces that people leave behind, both intentionally (e.g., as people contribute information for public consumption, whether in the form of a Facebook status update or a product review) and inadvertently (e.g., as people’s information seeking patterns are captured as they navigate from one location to another or query search engines), have allowed scholars to access a large, diverse, and enduring record of human behaviors, opinions, and interactions.

The Internet has additionally enabled large and geographically diverse communities to form and evolve, allowing scholars to examine the dynamic behaviors and

communication patterns of very large groups in situ. This has provided opportunities to track the diffusion of communication through large social networks that are difficult to access and observe offline, and has prompted new understanding of network dynamics on a very large scale (Bakshy, Rosenn, Marlow, & Adamic, 2012; Goel, Watts, & Goldstein, 2012; Sobkowicz, Kaschesky, & Bouchard, 2012). The emergence of Big Data and online communities has thus provided tremendous research opportunities that prompt both new variants on traditional methodological concerns of data sampling and integrity, as well as raise some new concerns that are largely specific to these data sources.

Methodological concerns and Big Data

As with more traditional research venues, two of the significant methodological concerns involving Big Data include sampling and data integrity. It can be challenging to apply traditional social science strategies and methods of sampling to Big Data. For example, defining units of analysis can be difficult because online behavioral or communication data can grow rapidly or change over time (Mazur, 2010; McAfee & Brynjolfsson, 2012). Additionally, some scholars have questioned who (i.e., what individuals) or what population is represented by Big Data, suggesting that because researchers lack knowledge about these individuals' characteristics, these data are often more limited in generalizability than conclusions may imply (Hargittai, 2015; Mahrt & Scharrow, 2013).

Although in many cases the proper sampling within these large datasets remains a critical methodological concern, it is also often the case that researchers can now conduct research on a *census* of behaviors or participants rather than a sample of them. Indeed, it is now common to analyze datasets where the number of observations is in the millions (Lin, 2015; McAfee & Brynjolfsson, 2012; Parks, 2014). Regarding data integrity, critical concerns with large datasets collected or observed online include variants of the threats to validity in traditional research (e.g., biases and social desirability effects), such as opinions that might be amplified or masked by the relative anonymity of participants (Wang et al., 2014) and ability to distinguish actual human behaviors from those initiated by bots or other automated response algorithms (Ratkiewicz et al., 2011; Stringhini, Kruegel, & Vigna, 2010), for example.

Many large datasets collected from online sources represent everyday behavior of people in their online environments, thus increasing ecological validity over laboratory experiments when research questions relevant to the online space are being explored (Mahrt & Scharrow, 2013). However, there are unique ethical considerations related to the use of Internet users' communication and behavioral traces. For example, researchers' observation of this behavior as well as collection of Internet users' data can occur without the usual informed consent required for most offline research (Buchanan, 2013; Eynon, Fry, & Schroeder, 2008). Furthermore, even when online interactions or communications are relatively public, Internet users may have varying perceptions of when and what use of their data for research entails a privacy violation.

Finally, because datasets drawn from online sources are often much larger than those traditionally used by social scientists, they sometimes require analytic tools that are

relatively novel to communication researchers (Agarwal & Dhar, 2014). While use of new tools to explore these datasets is an exciting future direction for communication research, analysis of particularly large datasets can potentially necessitate a trade-off between reliability and validity and statistical power (White & Breckenridge, 2014). Specifically, while large amounts of data are strong in power, the nature of the data may also limit the questions that social scientists can ask, as well as the measures that can be used in online social research (Mahrt & Scharkow, 2013). Thus, communication scholars and other social scientists can add to the study of Big Data by continuing to ask theory-driven research questions and by carefully considering the measures used to answer these questions.

Future directions: What communication scholars can contribute

The research affordances of new technology—including the availability of Big Data—provide communication scholars with a number of relatively novel research opportunities in scope and scale (see Parks, 2014). Indeed, there is currently a shift in the nature of communication research for those scholars interested in studying the online space, or in studying offline phenomena as informed by data available online. The growth of Big Data has increased the need for social scientists to be able to understand and apply computational or computer science techniques to social data (Welser, Smith, Fisher, & Gleave, 2008). To this increase in technical methods of studying online communities and interactions, communication scholars can bring their expertise in and theory-driven questions about the core of what is communicated and shared online, as well as their techniques for understanding human interaction and association.

For example, modeling of online communication networks is a growing and varied area of research that allows for both prediction of online behavior (e.g., Sobkowicz et al., 2012) as well as understanding of information diffusion through online social networks. Researchers in this area have examined the propagation of information via strong and weak social network ties (Bakshy et al., 2012), as well as the size and structure of online information cascades (Goel et al., 2012). However, the relationships indicated by these models of large amounts of online communications may be very different if the content of the communication between nodes is examined versus what might be assumed based solely on the structure of ties. For instance, some scholars question the extent to which online social network links between individuals are a valid indicator of interaction (e.g., Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009). Studying the content of communication between these links would help scholars not only understand when network links do represent interaction, but also the nature of those interactions.

Indeed, a significant *disadvantage* of contemporary interest in Big Data is that it sometimes can remove scholars further from the content of the communication inherent in much online data, as well as the social scientific theories that inform human understanding and behaviors (Snijders, Matzat, & Reips, 2012). To redress this shortcoming, communication scholars have the opportunity to bring theory-informed research questions regarding the nature of online communication to help to understand

patterns of online behavior and the large datasets that are currently drawing significant research interest across the increasingly allied disciplines of social, computer, and information science.

SEE ALSO: Big Data, Analysis of; Big Data, Collection of (Social Media, Harvesting); Online Research Methods, Qualitative; Quantitative Methodology; Sampling, Online; Social Network Analysis (Social Media); Survey Methods, Online; Unobtrusive Measures in Studying Social Media

References

- Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, 25(3), 443–448. doi:10.1287/isre.2014.0546
- Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519–528). New York: ACM Press. doi:10.1145/2187836.2187907
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368. doi:10.1093/pan/mpr057
- Buchanan, E. A. (2013). Internet research ethics: Past, present, and future. In M. Consalvo & C. Ess (Eds.), *The handbook of Internet studies* (pp. 83–108). Chichester, UK: Wiley-Blackwell.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3–5. doi:10.1177/1745691610393980
- Eynon, R., Fry, J., & Schroeder, R. (2008). The ethics of Internet research. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 23–41). Thousand Oaks, CA: SAGE.
- Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The structure of online diffusion networks. In *Proceedings of the 13th ACM conference on electronic commerce* (pp. 623–638). New York: ACM Press.
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76. doi:10.1177/0002716215570866
- Hughes, J. (2012). *SAGE Internet research methods*. Thousand Oaks, CA: SAGE.
- Lin, J. (2015). On building better mousetraps and understanding the human condition: Reflections on big data in the social sciences. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 33–47. doi:10.1177/0002716215569174
- Mahrt, M., & Scharrow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20–33. doi:10.1080/08838151.2012.761700
- Mazur, E. (2010). Collecting data from social networking web sites and blogs. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 77–90). Washington, DC: American Psychological Association.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 61–68. Retrieved from <https://hbr.org/2012/10/big-data-the-management-revolution> (accessed March 8, 2017).
- Palomares, N. A., & Flanagin, A. J. (2005). The potential of electronic communication and information technologies as research tools: Promise and perils for the future of communication research. In P. J. Kalbfleisch (Ed.), *Communication yearbook* (pp. 147–185). Mahwah, NJ: Lawrence Erlbaum Associates.

- Parks, M. R. (2014). Big data in communication research: Its contents and discontents. *Journal of Communication*, 64(2), 355–360. doi:10.1111/jcom.12090
- Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World Wide Web (WWW '11)* (pp. 249–252). doi:10.1145/1963192.1963301
- Reips, U. D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243–256. doi:10.1027//1618-3169.49.4.243
- Reips, U. D., & Krantz, J. H. (2010). Conducting true experiments on the web. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 193–216). Washington, DC: American Psychological Association.
- Scherpenzeel, A. C., & Bethlehem, J. G. (2011). How representative are online panels? Problems of coverage and selection and possible solutions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 105–132). New York: Routledge.
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology*, 57, 529–555. doi:10.1146/annurev.psych.57.102904.190048
- Smyth, J. D., & Pearson, J. E. (2011). Internet survey methods: A review of strengths, weaknesses, and innovations. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 11–44). New York: Routledge.
- Snijders, C., Matzat, U., & Reips, U. D. (2012). Big data: Big gaps of knowledge in the field of Internet science. *International Journal of Internet Science*, 7(1), 1–5.
- Sobkowicz, P., Kaschesky, M., & Bouchar, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4), 470–479. doi:10.1016/j.giq.2012.06.005
- Stringhini, G., Kruegel, C., & Vigna, G. (2010). Detecting spammers on social networks. *Annual Computer Security Applications Conference (ACSAC)*. doi:10.1145/1920261.1920263
- Tuten, T. L. (2010). Conducting online surveys. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 179–192). Washington, DC: American Psychological Association.
- Wang, G., Wang, B., Wang, T., Nika, A., Zheng, H., & Zhao, B. Y. (2014). Whispers in the dark: Analysis of an anonymous social network categories and subject descriptors. In *Proceedings of IMC '14*. New York: ACM Press. doi:10.1145/2663716.2663728
- Welser, H. T., Smith, M., Fisher, D., & Gleave, E. (2008). Distilling digital traces: Computational social science approaches to studying the Internet. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 116–140). Thousand Oaks, CA: SAGE.
- Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N., & Zhao, B. Y. (2009). User interactions in social networks and their implications. In *Proceedings of EuroSys '09*. New York: ACM Press. doi:10.1145/1519065.1519089
- White, P., & Breckenridge, R. S. (2014). Trade-offs, limitations, and promises of big data in social science research. *Review of Policy Research*, 31(4), 331–338. doi:10.1111/ropr.12078

Further reading

- Karpf, D. (2012). Social science research methods in Internet time. *Information, Communication & Society*, 15(5), 639–661. doi:10.1080/1369118X.2012.665468
- Keusch, F. (2012). How to increase response rates in list-based web survey samples. *Social Science Computer Review*, 30(3), 380–388. doi:10.1177/0894439311409709

- Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., Ruijven, T., ... & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, 45(3), 502–527. doi:10.1111/bjet.12067
- Peng, T. Q., Zhang, L., Zhong, Z. J., & Zhu, J. J. (2012). Mapping the landscape of Internet studies: Text mining of social science journal articles 2000–2009. *New Media & Society*, 15(5), 644–664. doi:10.1177/1461444812462846
- Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277–1291. doi:10.1007/s13278-012-0079-3
- Vehovar, V., & Manfreda, K. L. (2008). Overview: Online surveys. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 177–194). Thousand Oaks, CA: SAGE.

Kristin Page Hovevar is a doctoral candidate at the University of California at Santa Barbara. Her research focuses on information sharing using communication technologies and the social implications of the interactions, communities, and pooled information facilitated by these technologies. Specifically, she has examined motivations for sharing information via social media as well as the processes through which people evaluate and the judgments they make about information they find online.

Andrew Flanagin is a professor in the Department of Communication at the University of California at Santa Barbara. His research focuses on the ways in which information and communication technologies structure and extend human interaction, with particular emphasis on the processes of organizing and information evaluation and sharing.