Digital Traces: New Data, Resources, and Tools for Psychological-Science Research

Anat Rafaeli, Shelly Ashtar, and Daniel Altman

The William Davidson Faculty of Industrial Engineering & Management, Technion – Israel Institute of Technology

Abstract

New technologies create and archive *digital traces*—records of people's behavior—that can supplement and enrich psychological research. Digital traces offer psychological-science researchers novel, large-scale data (which reflect people's actual behaviors), rapidly collected and analyzed by new tools. We promote the integration of digital-traces data into psychological science, suggesting that it can enrich and overcome limitations of current research. In this article, we review helpful data sources, tools, and resources and discuss challenges associated with using digital traces in psychological research. Our review positions digital-traces research as complementary to traditional psychological-research methods and as offering the potential to enrich insights on human psychology.

As organizational psychologists, we research the influence of emotions expressed in a work-related context on various aspects of work. Our previous research in this area relied primarily on experimental manipulations or on self-reports of emotion (as reviewed by Hareli & Rafaeli, 2008; see Rafaeli et al., 2012, for an empirical example). More recently, we have been collaborating with computer-science colleagues, using automated tools to analyze data generated in online service conversations and study the effects of customer emotion on service agents. We analyzed 677,936 conversations to explore the evolution of customer emotion within conversations (G. B. Yom-Tov et al., 2018) and to test the effects of negative customer emotion on actual employee behavior. Our analysis demonstrated that employees respond more slowly (Altman, Ashtar, Olivares, & Yom-Tov, 2019) and take longer breaks (Ashtar, Yom-Tov, & Rafaeli, 2018) after interacting with customers expressing negative emotions. In another study, analyzing 8,259 conversations between customers and service agents, we showed that discrete emotions expressed by customers (e.g., anger) predict emotional behaviors of agents (Herzig et al., 2016). These recent studies—using large data samples representing actual behaviors of employees and customers—differ drastically from our previous lab- and self-report-based research and reveal the exciting opportunities that *digital traces* hold for psychological research. We review these opportunities in the hope of encouraging other psychological-science researchers to embrace them.

Technology—which increasingly mediates and supports human activities—retains digital traces of people's behaviors, creating a goldmine of data for psychological science. Digital traces can reveal people's intents, preferences, and emotions and also include aspects of the contexts in which actions occur (e.g., Stephens-Davidowitz & Pinker, 2017). Organizations use such data to define or assess business goals, and some analyses of this data are conducted as part of *computational social science* (Table 1; Alvarez, 2016). Yet their use in psychological research is still scant. Computer-science researchers are increasingly using tools common in their field to investigate topics more conventionally addressed by psychological scientists. Figure 1 documents, for example, the extensive growth of emotion research in computer science. However, the limited familiarity of researchers in this field with the theory and methodological rigor of psychological-science

Corresponding Author:
Anat Rafaeli, Technion – Israel Institute of Technology, The William Davidson Faculty of Industrial Engineering & Management, Bloomfield Building, Room 512, Technion City, Haifa 3200003, Israel
E-mail: anatr@ie.technion.ac.il

Table 1. Concepts, Definitions, and Examples From Digital-Traces Research

| Concept	Definition	Example	Resource
Computational social science	Use of computational approaches to model, simulate, and analyze social phenomena	Lazer et al. (2009) describe different types of data-driven studies in the field of computational social science.	Alvarez (2016) describes analytical methods for social research.
Digital-traces research	Research using data retained automatically by technological platforms	Stephens-Davidowitz and Pinker (2017) suggest social insights on the basis of different analyses of digital traces on Google.	Salganik (2018) provides a comprehensive review of tools and techniques for big-data research. Harlow and Oswald (2016) present a special issue on using big data in psychology. A Google tool for finding data sets can be found at https://toolbox.google.com/datasetsearch; https://www.kaggle.com is a platform for finding data sets.
Application programming interface (API)	A gateway for extraction of data using an interface of a specific digital platform	Jones, Wojcik, Sweeting, and Silver (2016) used the Twitter API to study emotion after violence on college campuses.	The Twitter API can be found at https://developer.twitter.com/. Murphy (2017) provides a guide to conducting psychological research on Twitter.
Web scraping	Automatic extraction of data from websites	E. Yom-Tov, Fernandez-Luque, Weber, and Crain (2012) scraped tags and text on Flickr to study pro-anorexia.	Landers, Brusso, Cavanaugh, and Collmus (2016) provide a primer on extraction of big data for psychological research; for related information, see https://rlanders.net/ and http://datascience.tntlab.org/.
Sensor data	Automatic recording of data collected with sensors	Lakens (2013) used smartphones to measure heart rate changes during relived happiness and anger.	Harari et al. (2016) review opportunities, practical considerations, and challenges of using smartphones to collect behavioral data. Mohr, Zhang, and Schueller (2017) review the use of ubiquitous sensors and machine learning for clinical-psychology research. Statista (2018) provides an overview of the Internet-of-things technology trends, innovations, security, and standardization issues.
Computer-aided text analysis (CATA)	Automatic coding of textual data	Speer (2018) used CATA to study narratives and performance evaluations.	Banks, Woznyj, Wesslen, and Ross (2018) review the use of CATA in psychology research and provide useful R tools.
Dictionary analysis	Automatic identification of categories in text by counting words defined in dictionaries	Pennebaker, Boyd, Jordan, and Blackburn (2015) describe the development and psychometric properties of Linguistic Inquiry and Word Count (LIWC), a dictionary-based CATA tool.	LIWC can be accessed at http://liwc.wpengine.com/.
Deep- or machine-learning text analysis	Automatic identification of topics in text using computational modeling	Speer (2018) used machine learning to analyze performance-evaluation texts.	A state-of-the-art Stanford tool for identifying emotion in text can be accessed at https://nlp.stanford.edu/sentiment/.
Sentiment analysis	Automatic identification and coding of positive and negative sentiment in text	Settanni and Marengo (2015) studied emotion and well-being through Facebook posts.	Cambria, Das, Bandyopadhyay, and Feraco (2017) provide a practical guide to sentiment analysis. An easy-to-use sentiment-analysis tool can be found at http://sentistrength.wlv.ac.uk/.
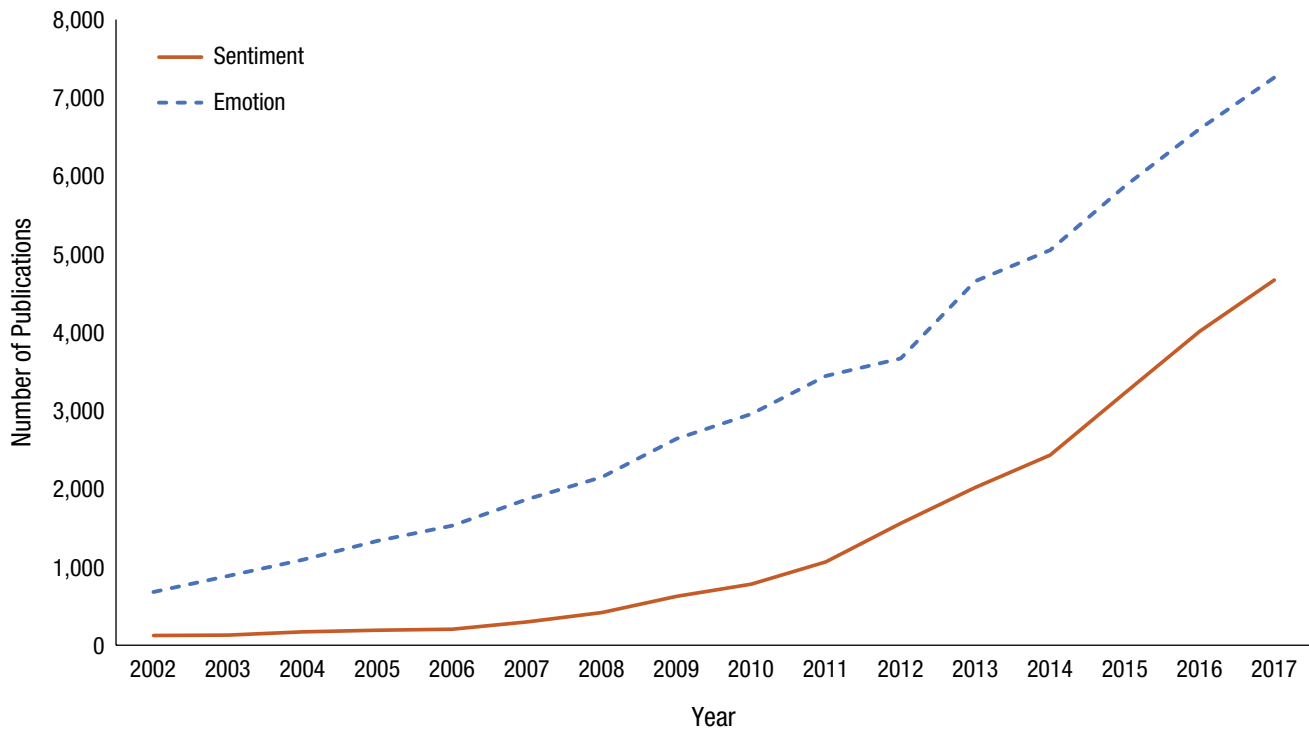
Fig. 1. Number of computer-science publications including the terms *sentiment* and *emotion* between 2002 and 2017. The figure illustrates substantial growth in the number of publications, suggesting that computer scientists are becoming more interested in psychological concepts. Data for the figure were extracted from the artificial-intelligence, image-processing, and information-systems categories on the Dimensions tool (https://www.dimensions.ai/).

research on emotion limits its potential depth. Furthermore, psychology researchers are not likely to review this research because of the unfamiliar concepts and methods it uses, as well as the unfamiliar journals in which it is published. Psychological scientists are currently less likely to use digital traces, which severely limits the potential breadth and impact of their research on this growing trend in computer science. We urge psychological scientists to step in and join digital-traces research both to enhance and to benefit from the versatility of this emerging field.

To this end, we describe how digital-traces data can enrich psychological research, review tools and resources for collection and analysis of digital-traces data, note useful hands-on guides for research with such data, and conclude with a review of challenges in such research.

Digital Traces: New Data for Psychological Research

Digital traces include records of website visits, product reviews, comments on social media, and more. Digital traces are collected and retained by Internet platforms, sensors, and other devices and typically comprise

contextual data about when, where, and for how long behaviors occurred. We suggest that there are three key merits of digital-traces data.

First, data can represent broader populations or target specific groups, extending studies beyond undergraduate students and Western societies. This is particularly useful with otherwise difficult-to-access groups that can be reached through designated websites (see https://support.therapytribe.com/). E. Yom-Tov, Fernandez-Luque, Weber, and Crain (2012), for example, studied pro-anorexia users using text and tags on the Flickr photo-sharing site (https://www.flickr.com/). A forum containing posts from people seeking an anorexic community (http://www.myproana.com/) can be useful for follow-up research. Similarly, researchers can use Twitter to access international populations. For example, 38.6 million active Twitter users are located in Japan. Indeed, 85% of the total 326 million Twitter users are located outside the United States. Online forums can facilitate access to professional communities (e.g., Stack Exchange: https://stackexchange.com/; Reddit nurses forum: https://www.reddit.com/r/Nurse/).

Second, digital traces provide fine-grained tracking of expressions and behaviors of large samples of

people. To illustrate, Twitter archives more than 500 million Tweets on an average day (http://www.internet livestats.com/twitter-statistics/), and over 4 million blog posts appear daily (http://www.worldometers.info/blogs/). These archives include information on both behavior (e.g., posts) and context (e.g., location). Sensors embedded in everyday objects (e.g., smart televisions) accrue a large volume of data (e.g., shows watched) on large numbers of participants (Greengard, 2015). The omnipresence of sensors, labeled the *Internet of things*, supports the accumulations of enormous amounts of data that can enable research on multiple and larger samples with less effort and resources than lab work or surveys.

Third, digital traces are automatic, unobtrusive records of digital expressions and behaviors that make up people's *digital dossier* (https://youtu.be/79IYZVYIVLA). Traces can be left intentionally (e.g., Facebook profiles) or unintentionally (e.g., details of mouse movement, https://www.clicktale.com/, or travel, https://www.google.com/maps/timeline). Digital traces can include activities, text, or photos and substantially reduce biases such as demand characteristics because people posting, tagging, or sharing photos or text online are unlikely to be aware of research goals.

These merits of digital-traces data can also facilitate comparison of results from different samples and verify reproducibility, a central issue for psychological science (Open Science Collaboration, 2015). As we discuss next, digital-traces research is supported by automated tools and resources that are rapidly evolving.

Digital Traces: New Tools and Resources for Psychological Research

Using digital traces in psychology research requires familiarization with new terminology and new tools for collecting and analyzing data. We briefly review must-know concepts and refer readers to comprehensive resources. Single rows in a data set—called *logs*—record an expression (e.g., a published text or picture) or a behavior (e.g., logging into a forum, heart rate) and include contextual data (e.g., time of action, location). Logs quickly accumulate into huge amounts of data of multiple types (e.g., textual, numeric, images, or videos), hence the term *big data*. Such data can be extracted from archives or collected through sensors. Publicly accessible archives include, for example, Wikipedia and Reddit, and there are search engines and platforms for finding data sets (e.g., https://toolbox.google.com/datasetsearch and https://www.kaggle.com). Digital-traces data can be collected through a data-collection interface, called an *application programming interface* (API; see Murphy, 2017, for a primer on API

retrieval of Twitter data) or directly extracted from websites in a process known as *web scraping*, which does not use APIs (see Landers, Brusso, Cavanaugh, & Collmus, 2016, for a web-scraping primer). Collaborating with organizations can also facilitate digital-traces research by providing access to intraorganizational data archives (G. B. Yom-Tov et al., 2018).

Sensors facilitate collection of fine-grained behavioral data (as opposed to surveys, experiments, or even diary studies) because they continuously document behaviors (Harari et al., 2016). Smartphones—which are today ubiquitous—can add another element to data collection, allowing communication with participants and collection of participant self-reports. To illustrate, Lathia, Sandstrom, Mascolo, and Rentfrow (2017) studied over 10,000 smartphone users and showed that physical activity (objectively measured with sensors) relates to (self-reported) happiness. Digital-traces data collected using sensors, smartphones, and wearable technology (e.g., Fitbit) free research from the constraints of labs and specific locations but require the complex translation of raw sensor data into meaningful indices of behavior and mental states (Mohr, Zhang, & Schueller, 2017). Matusik et al. (2019) described the use of wearable Bluetooth sensors for capturing relational variables and temporal variability in relationships. Lakens (2013) illustrated the use of sensors in an experimental paradigm by manipulating recalled emotion and measuring heart rate with a smartphone app.

Automated tools allow efficient analyses of large volumes of digital-traces data. Transcribing and coding voice and video can be done automatically (https://vi.microsoft.com/), reducing laborious research-assistant work. Written text can be analyzed using computer-aided text analysis (CATA), which relies on predefined dictionaries of terms. CATA can identify word clusters (Short, McKenny, & Reid, 2018) and topics (*topic modeling*; Banks, Woznyj, Wesslen, & Ross, 2018). The Linguistic Inquiry and Word Count (LIWC) tool described by Pennebaker, Boyd, Jordan, and Blackburn (2015) provides a word count of texts in predefined categories (e.g., counting words associated with power or with emotions). Reyt, Wiesenfeld, and Trope (2016), for example, used word counts to study the impact of high versus low construal level of advice givers on advice taking.

Dictionary analyses can be supplemented by incorporating grammatical structures into text analysis. Thelwall (2017), for example, used lexical supplements to separate texts such as "*not* angry" and "*very* angry," which would not be differentiated in a simple word-count analysis. State-of-the-art text analyses rely on *deep learning* (or other machine-learning methods), which use computations to train machines to automatically

code content. In these approaches, one sample of data trains a tool to classify content into categories. Trained tools are tested (validated) with other samples and allow coding of additional samples for further studies. Speer (2018), for example, used text analysis to derive narrative sentiment scores from qualitative performance evaluations in one sample and then applied these scores to an additional sample of (textual) performance data.

Sentiment analysis implements automated text analysis to study emotion (Cambria, Das, Bandyopadhyay, & Feraco, 2017). Herzig et al. (2016), for example, identified specific emotions (e.g., anger, frustration) of employees and customers. Settanni and Marengo (2015) used sentiment analysis to study emotion in Facebook posts. An additional implementation of text analysis is for automatic assessments of personality traits; Hinds and Joinson (2019) reviewed research in this domain. Table 1 presents brief explanations of concepts and useful examples and resources for digital-traces research.

Digital Traces: Challenges for Psychological-Science Research

We do not suggest digital traces as a replacement for current methods. Rather, digital traces can provide insights using more diverse, larger, and less biased data, as demonstrated above. Along with these opportunities, digital-traces research presents some challenges. First, the magnitude, redundancy, inaccuracies, and complexity of digital-traces data mean that raw data must be cleaned (sometimes referred to as *wrangling*; see Braun, Kuljanin, & DeShon, 2018), a process that can be extremely time consuming. Raw data often include duplicate records, typos, symbols or characters, and other "noise" that can distort even simple descriptive statistics, let alone inferential tests. Computing variables from raw data typically requires transforming data from its original form into a format that allows statistical analysis to address the research questions. The necessity for quality control of such transformations cannot be overemphasized; numbers are easy to produce and to compute, but the degree to which computed variables measure intended theoretical constructs is hardly straightforward. Speer (2018) illustrated this laborious process for CATA. Mohr et al. (2017) attempted to ease use of sensor data for research. Although a real challenge, the cleanup and quality control of transformations is rarely recognized or sufficiently thought through (Braun et al., 2018).

Second, the choice of platform from which data are obtained can create sampling biases (Ruths & Pfeffer, 2014) that require concerted attention. For example, Twitter data are attractive because they are relatively easy to retrieve (Murphy, 2017), but Twitter users tend to be millennials, college educated, and above-average income earners (Cooper, 2019). Forums and blogs represent a more diverse population, but retrieving their data is usually more difficult. Platforms can also pose challenges; Facebook, for example, is attractive as a research tool (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015), but its policies make retrieval of Facebook data difficult, perhaps impossible. Some researchers creatively overcome these policies: To illustrate, Settanni and Marengo (2015) asked participants to add the researcher as a Facebook friend and then collected information from participants' Facebook profiles.

Third, digital-traces data and research raise ethical concerns regarding privacy and informed consent. *Proceedings of the National Academy of Sciences, USA* (Verma, 2014), for example, posted an editorial concern about these issues following publication of Kramer, Guillory, and Hancock's (2014) study of emotion-contagion effects. Eliminating all identifying information (e.g., name, user ID) from collected data may not ensure participants' privacy and anonymity: Barbaro and Zeller (2006) identified a specific person despite removal of personally identifying information. People may not be aware of their participation in digital-traces research, and legal consent may mean long, obscure, and typically unread terms of use. Options to opt out are also somewhat obscure, so institutional review boards face an open dilemma as to whether ethical lines are crossed; the challenge is to balance the potential harm and potential benefit for social science.

Finally, obtaining digital-traces data requires skill and experience in programming and new statistical tools (e.g., Python, R) that are still not included in typical psychology curricula. However, resources for acquiring relevant knowledge are increasingly available and provide a viable path to capitalize on the opportunities that digital traces offer (Harlow & Oswald, 2016; Salganik, 2018; Table 1). Another path is collaborating with computer-science colleagues. For example, data on Internet platforms might not represent genuine human behavior; some data are placed maliciously by bots or hackers masquerading as legitimate users. This means that researchers must separate real-people data from nongenuine data, which is itself a challenge (Salge & Karahanna, 2018) but can be done with the help of computer scientists. Such collaborations can be challenging because of different disciplinary terms, methods, and motivations but also offer important interdisciplinary enrichment.

Once these and related challenges are overcome, a rich world of opportunities opens for psychologists. We hope our review is convincing in demonstrating that these challenges should not overrule the huge potential of digital-traces research. Social media platforms are

evolving, and new opportunities for insightful digital-traces studies surround us. Research of interest to psychology, for a better understanding of human behavior, should be conducted with full appreciation of psychological theories and research standards. We discovered digital-traces research when computer-science colleagues asked us for assistance with their research on emotion. We discovered a plethora of research on emotion being published in computer-science outlets, but for the most part, the data-driven nature of this research was not building on fundamental elements of psychological research; content and construct validity, reliability, and validity of measures and constructs, for example, are often missing. Psychology can help computer-science researchers create more theory-driven web scraping (Landers et al., 2016) as well as clarify variable definitions and hypothesized effects. Psychology can help itself by embracing digital-traces research and joining the big-data revolution.

Recommended Reading

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). (See References). Review of best-practices recommendations and tools for using text analysis in psychological research, including hypothesis and question formation, design and data collection, data preprocessing, and topic modeling; also discusses creation of scale scores for traditional correlation and regression analyses and provides an online repository for practice, an R markdown file, and an open-source topic-modeling tool.

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, *21*, 458–474. Clearly written, user-friendly, and relatively comprehensive guide for readers who wish to expand their knowledge on big-data research in psychology.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). (See References). Review of smartphone-sensing research highlighting opportunities for psychological research, considerations for designing studies, and methodological and ethical challenges.

Harlow, L. L., & Oswald, F. L. (Eds.). (2016). (See References). Special issue of *Psychological Methods* with various examples and perspectives on big-data research in psychology.

Salganik, M. J. (2018). (See References). Textbook with comprehensive coverage of concepts, tools, methods, and techniques for using big data and sources; includes discussion of data collection, surveys and experiments, and mass collaboration, as well as issues of ethics.

Action Editor

Randall W. Engle served as action editor for this article.

References

Altman, D., Ashtar, S., Olivares, M., & Yom-Tov, G. B. (2019). *Do customer emotions affect worker speed? An empirical study of emotional load in online customer contact centers*. Manuscript submitted for publication.

Alvarez, R. M. (Ed.). (2016). *Computational social science: Discovery and prediction*. Cambridge, England: Cambridge University Press.

Ashtar, S., Yom-Tov, G. B., & Rafaeli, A. (2018). *Demands and resources in the work of customer service agents: Predicting the length of unscheduled microbreaks* (Technical Report). Haifa, Israel: Technion – Israel Institute of Technology.

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*, 445–459.

Barbaro, M., & Zeller, T., Jr. (2006, August 9). A face is exposed for AOL Searcher No. 4417749. *The New York Times*. Retrieved from https://www.nytimes.com/2006/08/09/technology/09aol.html

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods*, *21*, 633–659.

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). *A practical guide to sentiment analysis* (Vol. 5). London, England: Springer.

Cooper, P. (2019). *28 Twitter statistics all marketers need to know in 2019*. Retrieved from https://blog.hootsuite.com/twitter-statistics/

Greengard, S. (2015). *The Internet of things*. Cambridge, MA: MIT Press.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*, 838–854.

Hareli, S., & Rafaeli, A. (2008). Emotion cycles: On the social influence of emotion in organizations. *Research in Organizational Behavior*, *28*, 35–59.

Harlow, L. L., & Oswald, F. L. (Eds.). (2016). Big data in psychology [Special issue]. *Psychological Methods*, *21*(4), 447–651.

Herzig, J., Feigenblat, G., Shmueli-Scheuer, M., Konopnicki, D., Rafaeli, A., Altman, D., & Spivak, D. (2016). Classifying emotions in customer support dialogues in social media. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 64–73).

Retrieved from http://www.aclweb.org/anthology/W16-3609

Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, *28*, 204–211.

Jones, N. M., Wojcik, S. P., Sweeting, J., & Silver, R. C. (2016). Tweeting negative emotion: An investigation of Twitter data in the aftermath of violence on college campuses. *Psychological Methods*, *21*, 526–541.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences. *American Psychologist*, *70*, 543–556.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, USA*, *111*, 8788–8790.

Lakens, D. (2013). Using a smartphone to measure heart rate changes during relived happiness and anger. *IEEE Transactions on Affective Computing*, *4*, 238–241.

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*, 475–492.

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLOS ONE*, *12*(1), Article e0160589. doi:10.1371/journal.pone.0160589

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., & Van Alstyne, M. (2009). Computational social science. *Science*, *323*, 721–723.

Matusik, J. G., Heidl, R., Hollenbeck, J. R., Yu, A., Lee, H. W., & Howe, M. (2019). Wearable Bluetooth sensors for capturing relational variables and temporal variability in relationships: A construct validation study. *Journal of Applied Psychology*, *104*, 357–387.

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, *13*, 23–47.

Murphy, S. C. (2017). A hands-on guide to conducting psychological research on Twitter. *Social Psychological & Personality Science*, *8*, 396–412.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943. doi:10.1126/science.aac4716

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Retrieved from https://repositories.lib.utexas.edu/handle/2152/31333

Rafaeli, A., Erez, A., Ravid, S., Derfler-Rozin, R., Efrat-Treister, D., & Scheyer, R. (2012). When customers exhibit verbal aggression, employees pay cognitive costs. *Journal of Applied Psychology*, *97*, 931–950.

Reyt, J., Wiesenfeld, B. M., & Trope, Y. (2016). Big picture is better: The social implications of construal level for advice taking. *Organizational Behavior and Human Decision Processes*, *135*, 22–31.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*, 1063–1064.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Salge, C. A. D. L., & Karahanna, E. (2018). Protesting corruption on Twitter: Is it a bot or is it a person? *Academy of Management Discoveries*, *4*, 32–49.

Settanni, M., & Marengo, D. (2015). Sharing feelings online: Studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology*, *6*, Article 1045. doi:10.3389/fpsyg.2015.01045

Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, *5*, 415–435.

Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, *71*, 299–333.

Statista. (2018). *IoT market and technology trends worldwide 2018*. Retrieved from https://www.statista.com/study/53546/iot-market-and-technology-trends-worldwide-2018/

Stephens-Davidowitz, S., & Pinker, S. (2017). *Everybody lies: Big data, new data, and what the Internet can tell us about who we really are*. New York, NY: HarperCollins.

Thelwall, M. (2017). Heart and soul: Sentiment strength detection in the social web with SentiStrength. In J. Holyst (Ed.), *Cyberemotions: Collective emotions in cyberspace* (pp. 119–134). Berlin, Germany: Springer.

Verma, I. M. (2014). Editorial expression of concern and correction. *Proceedings of the National Academy of Sciences, USA*, *111*, 10779.

Yom-Tov, E., Fernandez-Luque, L., Weber, I., & Crain, S. P. (2012). Pro-anorexia and pro-recovery photo sharing: A tale of two warring tribes. *Journal of Medical Internet Research*, *14*(6), Article e151. doi:10.2196/jmir.2239

Yom-Tov, G. B., Ashtar, S., Altman, D., Natapov, M., Barkay, N., Westphal, M., & Rafaeli, A. (2018). *Customer sentiment in web-based service interactions: Automated analyses and new insights* (pp. 1689–1697). In WWW '18 Companion Proceedings of The Web Conference 2018. Lyon, France: International World Wide Web Conference Committee.