Digital Traces:

New Data, Resources and Tools for Psychological Science Research

Anat Rafaeli [i], Shelly Ashtar, Daniel Altman

Technion- Israel Institute of Technology

Current Directions in Psychological Science

First Published August 14, 2019 | Research Article

https://doi.org/10.1177/0963721419861410

Abstract

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

Keywords: Digital Traces; Big Data; Automated Data Collection; Computer-Aided Text Analysis; Sentiment Analysis.

Main Text

As organizational psychologists, our group studies 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, Rafaeli, & Yom-Tov, 2017) 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/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 imply people's intents, preferences and emotions, and also include aspects of the context (e.g., when actions occur; Stephens-Davidowitz & Pinker, 2017). Organizations use such data to define or assess business goals, and some analyses of such data are conducted as part of "Computational Social Science" (Table 1; Alvarez, 2016), yet 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 this field with the theory and methodological rigor of psychological science research on emotion limits the potential depth of this research. Furthermore, psychology researchers are not likely to review this research, because of the unfamiliar journals, concepts and methods it uses. 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.

We thus 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.

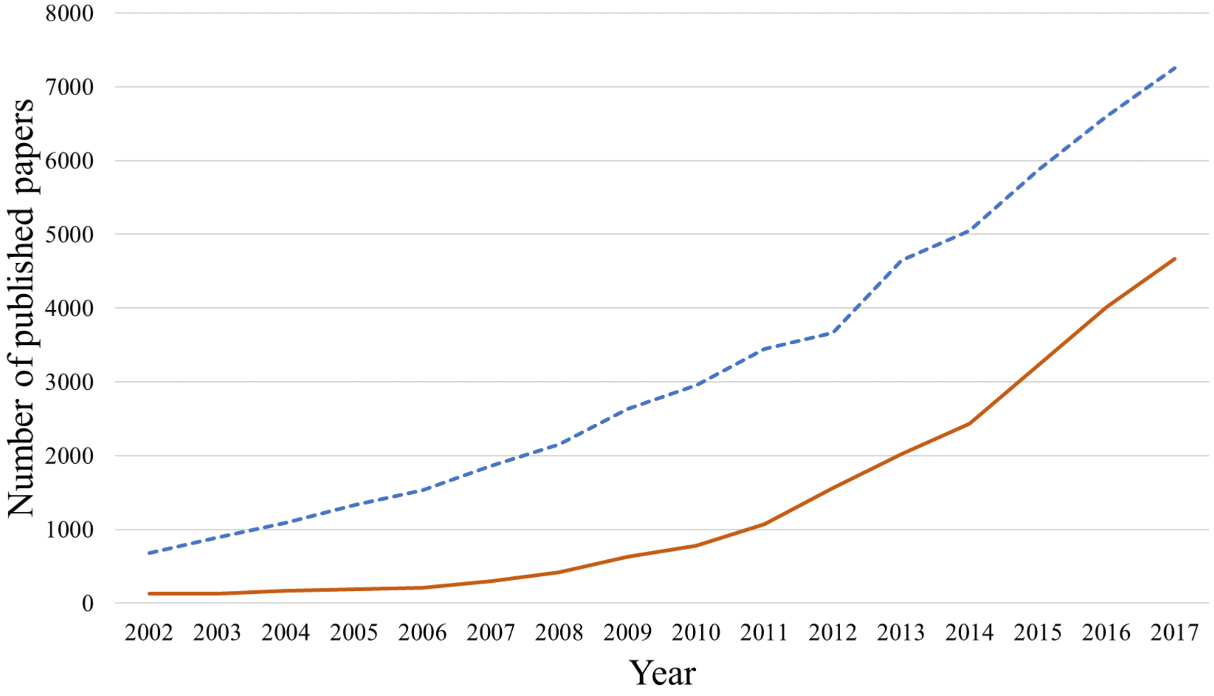

Figure 1. Number of published Computer Science papers including the terms “sentiment” and “emotion” between 2002 and 2017. The figure indicates substantial growth in number of papers published, suggesting interest of Computer Scientists in psychological concepts. Data for the figure are extracted using https://www.dimensions.ai/, regarding research on: (i) Artificial Intelligence and Image Processing, and (ii) Information Systems.

DIGITAL TRACES: NEW DATA FOR PSYCHOLOGICAL RESEARCH

Digital Traces include records of website visits, product reviews, and comments in 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 three key merits of Digital Traces data:

(i) 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 (cf. https://support.therapytribe.com/). E. Yom-Tov, Fernandez-Luque, Weber, and Crain (2012), for example, studied pro-anorexia users using text and tags in the Flickr photo-sharing site (https://www.flickr.com/). A forum including 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 US. Online forums can facilitate access to professional communities (cf. Stack Exchange https://stackexchange.com/; Reddit nurses forum https://www.reddit.com/r/Nurse/).

(ii) 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.internetlivestats.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 which can enable research on multiple and larger samples with less effort and resources than lab work or surveys.

(iii) 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 cognizantly (e.g., Facebook), 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 photo (e.g., https://www.flickr.com/), and substantially reduce biases such as demand characteristics, since 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 dataset --called ***logs***-- record an expression (e.g., a published text, or picture), and/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 multiple-type data (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 datasets (cf.

https://toolbox.google.com/datasetsearch and https://www.kaggle.com). Digital Traces data can be collected through a data collection interface, called an API (***Application Programming Interface***; cf. Murphy, 2017 primer for API retrieval of Twitter data) or directly extracted from websites in a process known as ***Web Scraping*** which does not use APIs (cf. Landers, Brusso, Cavanaugh, & Collmus, 2016 primer). Collaborating with organizations can also facilitate Digital Traces research, by providing access to intra-organizational data archives (cf. G. B. Yom-Tov et al., 2018).

Sensors facilitate collection of fine-grained behavioral data (as opposed to surveys, experiments or even diary studies), since they continuously document behaviors (cf. 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 wearables (e.g., Fitbit) frees research from the constraints of labs and specific locations, but requires the complex translation of raw sensor data into meaningful indices of behavior and mental states (Mohr, Zhang, & Schueller, 2017). Matusik et al. (2018) describe the use of wearable bluetooth sensors for capturing relational variables and temporal variability in relationships. Lakens (2013) illustrates 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 (cf. https://vi.microsoft.com/), reducing laborious research assistant work. Written text can be analyzed using ***Computer Aided Text Analysis (CATA),*** which relies on pre-defined ***dictionaries*** of terms. CATA can identify word clusters (Short, McKenny, & Reid, 2018) and topics (***topic modeling***; cf. 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, emotions, etc.). Reyt, Wiesenfeld, and Trope (2016), for example, used word counts to study the impact of (high vs. low) c*onstrual 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 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 (cf. 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) review research in this domain.

Table 1 presents brief explanations of concepts and useful examples and resources for Digital Traces research.

Table 1

Brief explanations of concepts and useful examples and resources for Digital Traces research

Concepts		Brief Explanation, Examples and Resources
Computational Social Science	Brief Explanation	Using computational approaches to model, simulate, and analyze social phenomena.
	Example	Lazer et al. (2009): Describe different types of data-driven “Computational Social Science” studies.
	Resource	Alvarez (2016): Provides analytical methods for social research.
Digital Traces Research	Brief Explanation	Research using data retained automatically by technological platforms.
	Example	Stephens-Davidowitz and Pinker (2017): Suggest social insights based on different analyses of Digital Traces on Google.
	Resources	Salganik (2017): Comprehensive review of tools and techniques for Big Data research.
		Harlow and Oswald (2016): Special issue on using Big Data in psychology
		Google tool for finding datasets: https://toolbox.google.com/datasetsearch
		A platform for finding data sets https://www.kaggle.com
Application Programming Interface (API)	Brief Explanation	A gateway for extraction of data using an interface of a specific digital platform.
	Example	Jones et al. (2016): Used Twitter API to study emotion after violence on college campuses.
	Resources	Twitter API https://developer.twitter.com/
		Murphy (2017): A guide to conducting psychological research on Twitter.
Web Scraping	Brief Explanation	Automatic extraction of data from websites.
	Example	E. Yom-Tov et al. (2012): Used scraping of tags and text on Flickr to study pro-anorexia.
	Resource	Landers et al. (2016): Primer on extraction of Big Data from the internet for psychological research
Sensor Data	Brief Explanation	Automatic recording of data collected with sensors.
	Example	Lakens (2013): Used Smartphones to measure heart rate changes during relived happiness and anger.
	Resources	Harari et al. (2016): Review of opportunities, practical considerations, and challenges of using

		Smartphones to collect behavioral data. Mohr et al. (2017): Review of using ubiquitous sensors and machine learning for clinical psychology research. https://www.statista.com/study/53546/iot-market-and-technology-trends-worldwide-2018/: Overview of the Internet of Things (IoT) technology trends, innovations, security and standardization issues.
Computer-Aided Text Analysis (CATA)	Brief Explanation	Automatic coding of textual data.
	Example	Speer (2018): Used Computer-Aided Text Analysis to study narratives performance evaluations.
	Resource	Banks, Woznyj, Wesslen, and Ross (2018): Review of using Computer Aided Text Analysis in psychology research and provide useful R tools.
Dictionary Analysis	Brief Explanation	Automatic identification of categories in text by counting words defined in dictionaries.
	Example	Pennebaker, Boyd, Jordan, and Blackburn (2015): describes the development and psychometric properties of LIWC, a dictionary-based computer aided text analysis tool.
	Resource	http://liwc.wpengine.com/: A tool for analyzing texts using dictionary analysis.
Deep or Machine Learning Text Analysis	Brief Explanation	Automatic identification of topics in text using computational modelling.
	Example	Speer (2018): Use machine learning to analyze performance evaluation texts.
	Resource	https://nlp.stanford.edu/sentiment/: State-of-the-art "Stanford" tool for identifying emotion in text.
Sentiment Analysis	Brief Explanation	Automatic identification and coding of positive and negative sentiment in text.
	Example	Settanni and Marengo (2015): Study emotion and well-being through Facebook posts.
	Resource	Cambria et al. (2017): A practical guide to Sentiment Analysis; http://sentistrength.wlv.ac.uk/: An easy to use Sentiment Analysis tool.

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. Together 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” (cf. Braun, Kuljanin, & DeShon, 2018)), a process that can be extremely time consuming. Raw data often includes 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) illustrates this laborious process for Computer-Aided Text Analysis. Mohr et al. (2017) attempt 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 is obtained can create sampling biases (Ruths & Pfeffer, 2014) that require concerted attention. For example, Twitter data is attractive because it is relatively easy to retrieve (Murphy, 2017), but Twitter users tend to be Millennials, college educated and earn above average incomes (https://blog.hootsuite.com/twitter-statistics/). 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. PNAS, for example, posted an editorial concern about these issues following publication of Kramer, Guillory, and Hancock's (2014) study of emotion contagion effects (http://www.pnas.org/content/pnas/111/29/10779.1.full.pdf). 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 IRB committees face an open dilemma as to whether ethical lines are crossed; the challenge is to balance between the potential harm and 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, providing a viable path to capitalize on the opportunities that Digital Traces offer (cf. Harlow & Oswald, 2016; Salganik, 2017; Table 1). Another path is collaborating with Computer Science colleagues. For example, data on internet platforms might *not* represent genuine human behavior; some data is placed maliciously, by bots or hackers masquerading as legitimate users. This means that researchers must separate "real-people" data from non-genuine 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 standard. 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.

Acknowledgments and Endnote

We thank Rakefet Ackerman, Ofra Amir, Avigdor Gal, Ailie Marx, Alina Shaulov and four anonymous reviewers for their insightful comments on an earlier version of this paper. Work on the paper was partially funded by the generous support of Israel's National Institute of Health Policy (Grant #702899) and the Israel Science Foundation (Grant #1717/13).

i Address correspondence to Anat Rafaeli, anatr@ie.technion.ac.il

References

Altman, D., Rafaeli, A., & Yom-Tov, G. B. (2017). When psychology meets operations: the influence of cognitive and emotional loads on the efficiency of customer-service employees. In *INFORMS annual meeting*. Houston, Texas.

Alvarez, R. M. (2016). *Computational Social Science*. 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. *Manuscript in Preparation*.

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*(4), 445–459.

Barbaro, M., & Zeller, T. (2006). A face is exposed for AOL searcher No. 4417749. 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*(3), 633–659.

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

Greengard, S. (2015). *The internet of things*. 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*(6), 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. (2016). Big Data in Psychology [Special issue]. *Psychological Methods*, *21*(4).

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. *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, (September), 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*, 1–8.

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*(4), 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*(6), 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*, *111*(24), 8788–8790.

Lakens, D. (2013). Using a smartphone to measure heart rate changes during relived happiness and anger. *IEEE Transactions on Affective Computing*, *4*(2), 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*(4), 475.

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), 1–13.

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

Matusik, J. G., Heidl, R., Hollenbeck, J. R., Yu, A., Lee, H. W., & Howe, M. (2018). Wearable bluetooth sensors for capturing relational variables and temporal variability in relationships: A construct validation study. *Journal of Applied Psychology*.

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 and Personality Science, 8*(4), 396–412.

Open Science Collaboration. (2015). *Science, 349*, aac4716. https://doi.org/10.1126/science.aac4716

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.

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*(5), 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*(6213), 1063–1064.

Salganik, M. J. (2017). *Bit by bit: social research in the digital age*. 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*(1), 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*(JUL), 1–7.

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.

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: HarperCollins.

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

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), e151.

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. In *WWW '18 Companion: The 2018 Web Conference Companion*. Lyon, France.

Recommended Readings

1. Clearly written, user-friendly, and relatively comprehensive guide for readers who wish to expand their knowledge on Big Data research in psychology.

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to Big Data research in psychology. *Psychological Methods*, *21*(4), 458–474.

2. Textbook with comprehensive coverage of concepts, tools, methods and techniques for using Big Data techniques and sources. Includes discussion of data-collection, surveys and experiments, and mass collaboration, as well as issues of ethics.

Salganik, M. J. (2017). *Bit by bit: social research in the digital age*. Princeton University Press.

3. Special issue of Psychological Science with various examples and perspectives on Big Data research in psychology.

Harlow, L. L., & Oswald, F. L. (Eds.) (2016). Big Data in Psychology [Special issue]. *Psychological Methods*, *21*(4).

4. Review, best practice recommendations and tools for utilizing text analysis in psychological research, including hypothesis and question formation, design and data collection, data pre-processing, 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 modelling tool.

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

.

5. Review of Smartphone-sensing research highlighting opportunities for psychological research, considerations for designing studies, and methodological and ethical challenges.

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*(6), 838–854.