



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Archival research: Expanding the methodological toolkit in social psychology

Yu Tse Heng^{a,*}, David T. Wagner^b, Christopher M. Barnes^c, Cristiano L. Guarana^d^a University of Washington, 359 Mackenzie, Seattle, WA 98195, United States^b University of Oregon, 473 Lillis, Eugene, OR 97405, United States^c University of Washington, 585 Paccar, Seattle, WA 98195, United States^d Indiana University, 1309 E. Tenth Street, Bloomington, IN 47405, United States

ARTICLE INFO

Handling editor: Roger Giner-Sorolla

Keywords:

Archival research
Research methods
Big Data
Social psychology

ABSTRACT

Laboratory experiments have many benefits and serve as a powerful tool for social psychology research. However, relying too heavily on laboratory experiments leaves the entire discipline of social psychology vulnerable to the inherent limitations of laboratory research. We discuss the benefits of integrating archival research into the portfolio of tools for conducting social psychological research. Using four published examples, we discuss the benefits and limitations of conducting archival research. We also provide suggestions on how social psychological researchers can take advantage of the benefits while overcoming the weaknesses of archival research. Finally, we provide useful resources and directions for utilizing archival data. We encourage social psychologists to increase the robustness of this scientific literature by supplementing laboratory experiments with archival research.

Social psychology has a long and respected tradition of conducting laboratory experiments. There are clear benefits to conducting such experiments. Most notably, laboratory experiments include the elements of contextual control and random assignment to treatment and control groups that when utilized properly allow researchers to draw causal inferences (Cook & Campbell, 1979). Delineation of causality allows for the generation and refinement of psychological theories, and aids in the understanding of how to influence psychological phenomena. Laboratory experiments are tailor made to facilitate these inferences, making them an extremely powerful and useful tool for conducting social psychological research (Falk & Heckman, 2009).

Despite their many strengths, laboratory experiments have important limitations. Artificial settings may miss important elements of real world contexts (Kerlinger, 1986), and demand characteristics in such artificial settings can distort construct relationships (Klein et al., 2012). Laboratory experiments are often conducted with relatively small samples, which may lead to unstable parameter estimates and invalid inferences (Hollenbeck, DeRue, & Mannor, 2006), and undermine the reliability of replications (Fraleigh & Vazire, 2014; Open Science Collaboration, 2015). Some of these limitations contribute to what some are calling a “crisis of confidence” in psychology (Baumeister,

2016; Hales, 2016; Pashler & Wagenmakers, 2012). Also, impracticality of random assignment of some characteristics potentially narrows the range of topics that can be studied in laboratory experiments (Doss, Rhoades, Stanley, & Markman, 2009; Sbarra, Emery, Beam, & Ocker, 2014).

Archival research has the potential to address many of these limitations and is therefore a promising complementary research approach to the traditional laboratory experiments. Archival research entails analyzing data that were stored other than for academic research purposes¹. This research approach has frequently been utilized in other fields (e.g., economics, sociology, and developmental psychology; Cherlin, 1991; Shultz, Hoffman, & Reiter-Palmon, 2005), but remains severely underutilized in social psychology. A search of the published articles in three top social-psychology journals (Journal of Personality and Social Psychology, Psychological Science, and Journal of Experimental Social Psychology) in 1996, 2006, and 2016 reveals that archival studies were used in < 1% of the published studies across three decades, meaning that only a small subset of the social psychology literature uses archival research. This underrepresentation of archival research is evident in spite of the high-impact archival studies that have been done in the field, such as Cohn, Mehl, and Pennebaker's (2004)

* Corresponding author.

E-mail addresses: yutse@uw.edu (Y.T. Heng), dwagner@uoregon.edu (D.T. Wagner), Chris24b@uw.edu (C.M. Barnes), cguarana@iu.edu (C.L. Guarana).¹ Although meta-analyses can be categorized as archival (e.g., Scandura & Williams, 2000), we consider this to be a different type of research that is already well utilized by social psychologists. We also exclude from our definition large-scale pre-existing datasets that were collected for academic purposes (e.g., American Time Use Survey, Studies from the GLOBE project, World Values Survey) even though these datasets are archival in nature and may offer triangulating value to researchers.

Table 1
Summary of archival research case example characteristics.

Archival research case example	Research design	Archival data sample size	Type of measures	Archival data availability	Combined studies
The Facebook A/B Study	True experiment	> 3 million Facebook posts	Facebook posts	Not publicly accessible	No
The Sleep and Cyberloafing Study	Natural experiment	3492 searches	Google trends daily search volume	Publicly accessible	Yes – observational lab study
The Divorce Education Program Study	Quasi-experiment	434 families	Divorce decrees and parenting plans	Publicly accessible	No
The Anticipatory Consumption Study	Correlational study	149 newspaper articles	Newspaper archives	Publicly accessible	Yes – survey, experience sampling study, and experiment

study on linguistic markers of psychological change after the September 11 attacks, Alter and Oppenheimer's research (2008) on the effects of fluency, and Sales' (1973) investigation of threat as a cause of authoritarianism.

Considering that the digital universe will more than double every two years from 2013 to 2020—from 4.4 trillion to 44 trillion gigabytes (International Data Corporation, 2014), archival research can be a fruitful and robust methodology for social psychologists to investigate social phenomena. Yet despite the vast amount of data available, only half of 1% of newly created digital data have been analyzed (MIT Technology Review, 2013). In recent years, tools for the assembly of relevant datasets have become widely available to researchers, including notable examples such as Google Trends, Twitter tags, and online marketplace bidding logs. Clearly, the “Big Data” revolution is beginning to alter the research landscape by turning archival research into a promising methodological option for research.

Archival research can take many forms, including true experiments, natural experiments, quasi-experiments, and correlational studies. Such data tend to occur in natural social settings, which offer social psychologists the opportunity to directly examine real-world phenomena that, by comparison, are often artificially simulated in laboratory settings. The massive and diverse samples typical of archival studies also yield several benefits, such as increased statistical power and generalizability. However, features of archival data have drawbacks that could result in researchers drawing misleading conclusions based upon null-hypothesis significance tests attached to small effect sizes, and the introduction of other forms of biases. Also, it is worth noting that even archival approaches to research call for a consideration of unsound research practices pertaining to data collection, measurement validity, and ethical concerns. As such, archival research has the potential to increase the robustness of social psychology research, but researchers need to be mindful of the potential limitations that accompany such an approach.

In this paper, we contrast the pros and cons of archival research by its key features (nature of data, sample characteristics, and type of measures) to assess its added value to archival social psychological researchers. We draw from four archival research case examples that respectively adopt a true experiment, natural experiment, quasi-experiment, and correlational research design to illustrate these strengths and weaknesses. We also suggest potential solutions to these weaknesses. These include additional recommendations to deal with open practices concerns specific to archival research and steps that researchers can take to reduce data reliability and validity concerns. Finally, we provide researchers with starting points and directions to conducting archival research (e.g., available resources for data acquisition/processing, useful statistical techniques, and novel archival research approaches).

1. Four archival research case examples

We discuss the strengths and weaknesses of the various features of archival research with respect to four recent papers that utilized an

archival research approach, which we refer to as *The Facebook A/B Study* (Kramer, Guillory, & Hancock, 2014), *The Sleep and Cyberloafing Study* (Wagner, Barnes, Lim, & Ferris, 2012), *The Divorce Education Program Study* (deLusé & Braver, 2015), and *The Anticipatory Consumption Study* (Kumar, Killingsworth, & Gilovich, 2014). We supplement these four examples with additional archival studies. The four studies were specifically chosen to highlight the range of research designs in archival research. Whereas most researchers are familiar with true experiments and correlational studies, the terms “natural experiment” and “quasi-experiment” are often used interchangeably and loosely in the literature. To be precise, in natural experiments, the treatment is a result of a naturally occurring or unplanned event that was not intended to influence the outcome of interest. On the other hand, in quasi-experiments, the treatment is planned and resembles a randomized experiment but lacks a full random assignment (Remler & Ryzin, 2015). Correlational studies are the most predominant research method in archival studies. In the domain of published archival research, there are only very few true experiments, natural experiments, and quasi-experiments². Plausible explanations for the lack of such archival experimental designs include the rarity of opportunities to introduce a manipulation into the real world, the rarity of a serendipitous occurrence of an unplanned event that is relevant to the social psychologist's research, and the rarity of planned treatments that are retrospective (it is more common for prospective data to be collected when treatments are intentionally introduced).

Yet despite rare evidence of such studies in the literature, the opening discussion of this paper highlights the ubiquity and generation of seemingly infinite amounts of data. In the midst of such munificence, substantial insight can be gained by researchers who are willing to broaden their conception of what constitutes social psychological research. To this end, this paper makes a case for archival research as a propellant for our field, and to a certain extent, as a remedy for some of the maladies that ail the field. We summarize the four archival research case examples below and in Table 1.

1.1. The Facebook A/B study (true experiment)³

Kramer et al. (2014) demonstrated in a study on Facebook users that when positive content on Facebook feeds was reduced, people produced

² There is a particularly low likelihood of archival true experiments and quasi-experiments as researchers who artificially create procedures to manipulate conditions would likely use them for prospective research. However, we still include these types of research in order to provide a complete range of archival research options.

³ As pointed out by an anonymous reviewer, the prospective nature of introducing a manipulation in the Facebook A/B Study may seem misaligned with the definition of archival research. However, given that the authors indicated in their journal submission that the “experiment was conducted by Facebook, Inc. for internal purposes” (Verma, 2014, p. 10779), suggesting that publishing the data for academic purposes was ancillary to the original commercial aim of the experimentally generated data, the Facebook A/B Study fits the definition of an archival research study. If the data had instead been collected by the research team with the primary purpose of generating publishable research, then the study would be considered a true experiment or a field experiment, but not be considered archival research.

fewer positive and more negative Facebook posts, and the opposite occurred when negative content on Facebook feeds was reduced. The researchers obtained data from an internal market research A/B test conducted by Facebook. A common term in marketing, A/B testing is a basic randomized controlled experiment where two versions of a variable (e.g., an ad, email, or web page) are tested to see which version performs better (Gallo, 2017). For their study, Facebook manipulated the amount of positive and negative content in Facebook newsfeeds. This example highlights how procedures like A/B testing can potentially be a source of true archival experiments.

1.2. The sleep and cyberloafing study (natural experiment)

Wagner et al. (2012, Study 1) showed that lost sleep increased cyberloafing. In a natural experiment, the researchers demonstrated the effects of lost sleep by comparing cyberloafing levels on the Monday after the shift to Daylight Saving Time with cyberloafing levels on the Mondays immediate preceding and following this day. To measure cyberloafing, the researchers obtained internet search volumes from the Google Trends database, collating these logs from over 200 of the largest U.S. metropolitan areas for the years 2004 to 2009. Cyberloafing was operationalized as the relative percentage of internet searches conducted in the “Entertainment” category (e.g., YouTube, ESPN, and videos) on a given day, with the assumption that these websites are unrelated to work. They also tested their predictions in a controlled laboratory study.

1.3. The divorce education program study (quasi-experiment)

deLusé and Braver (2015) found that attending a single two-hour divorcing parent education class resulted in an increase in the amount of visitation time awarded. Using a quasi-experimental design, the researchers compared divorce outcomes of divorced families that attended (treatment group) and did not attend (control group) the divorce education program. These groups were not randomly assigned—the treatment group had petitioned for divorce within a specific six-week period where the judge had ordered them to attend the class, while the control group had petitioned for divorce six weeks immediately before or after the treatment interval. The researchers obtained the visitation time measure from the archived divorce decrees and parenting plans of these divorced families.

1.4. The anticipatory consumption study (correlational study)

Kumar et al. (2014, Study 3) found that people tend to feel more positive when they are waiting for experiences than when they are waiting for possessions. Adopting a correlational study design, the researchers analyzed newspaper articles about people waiting in line to make a purchase. The researchers collated these articles by searching the LexisNexis database over a two-year period, using search terms such as “line AND wait” and “wait AND hours”. To measure the type of purchase, coders rated the extent to which whatever the individuals were waiting for was experiential or material. Similarly, to measure mood, coders rated these individuals on how positive or negative their mood or behavior seemed. The researchers complemented this archival study with other studies, including a survey, an experience-sampling study, and a randomized experiment.

2. Features of archival research

2.1. Nature of data

2.1.1. Uses pre-existing datasets

A key feature that distinguishes archival research from traditional research is the use of pre-existing data in the former and the use of prospective data in the latter. The traditional research process follows

the order of research question generation, research design development, data collection, and then data analysis. However, the archival research process order is different in that data collection comes before the other research steps. This difference poses some challenges to conducting archival research in line with the open science best practices recently put forth in psychology (Nosek, Ebersole, DeHaven, & Mellor, 2018; van't Veer & Giner-Sorolla, 2016).

Social psychology as a field is warming up to the call for more openness, transparency, and reproducibility in our research process (for comprehensive open science resources, see: <http://cos.io> and <http://osf.io>). The Transparency and Openness Promotion (TOP) guidelines serve as a crucial milestone toward an open research culture, providing concrete strategies and recommendations for researcher and publishers on citation standards, transparency (data, analytic methods/code, research materials, design, and analysis), pre-registration (study, analysis plan), and replication (Nosek et al., 2015; for an excellent summary, refer to: <https://cos.io/our-services/top-guidelines/>). In the following discussion, we build on the work of the TOP committee, the Open Science Framework, and Nosek et al. (2018) to address openness, transparency, and reproducibility concerns specific to archival research.

We first discuss the need for and challenges associated with pre-registering archival research. Growing concerns about unethical research practices such as p-hacking, HARK-ing (hypothesizing after the results are known, Kerr, 1998), and cherry-picking have led researchers to become more interested in the pre-registration of research. In pre-registration, researchers describe their hypotheses, methods, and analyses prior to conducting the study, in a manner that can be externally verified (van't Veer & Giner-Sorolla, 2016). Ideally, pre-registration is a solution to these unethical research practices, as publicly timestamping study plans and predictions would mean that a researcher cannot claim that a post-hoc hypothesis was decided before data were collected. However, in archival research, the availability of information about the archival data (albeit to varying extents) prior to the generation of research questions makes it difficult to ensure any “pure” pre-registration of archival research. This familiarity with the contents of the data could lead to biases that pre-registration is intended to prevent, even in well-intentioned researchers. For instance, biases may be formed when a researcher reads a published paper that was based on the archival dataset, even if the focal variables are not strongly related to the researcher's study.

We provide some recommendations to minimize these concerns about the pre-registration of archival studies. First, we suggest that researchers report whether the archival dataset was identified prior to research question generation. If the archival dataset was identified beforehand, researchers should transparently and clearly report the extent of pre-known information about the dataset prior to research question generation (e.g., any preliminary data analyses that were conducted, knowledge of any known empirical findings from the same dataset, such as from published studies). Second, we encourage researchers to report the process of setting the scope of their research question, explaining clearly any iterations between initial data exploration and research question generation. One way to reduce the likelihood of researcher bias is to clearly distinguish the research question generator and data analyst roles within the research team itself, such that the research question generators are not involved in data analysis, and data analysts are not involved in the generation of the research questions. We strongly recommend that researchers include these details in their study pre-registrations. These important steps will certify that the researchers either adopted a confirmatory approach to hypothesis testing, thereby increasing confidence in research findings, or honestly reported their exploratory approach, which is also acceptable as long as research conclusions are appropriately limited.

Next, we discuss the concerns associated with archival data transparency, specifically (1) data acquisition/construction transparency, and (2) accessibility of data. In general, researchers obtain a pre-

existing dataset either in its entirety or construct the dataset by web scraping or using APIs. Given the rarity of archival research in social psychology, existing open science guidelines insufficiently address the need for transparency about the dataset construction procedure. Especially in the case of constructed archival datasets, we recommend that researchers report in detail their data acquisition process, such as the web scraping script, programming language used, persons involved in the dataset construction and what their roles were, and the checks and balances that were introduced to prevent technical errors. Researchers should also report decisions about data transformation and corrections, how missing values were treated, as well as data inclusion or exclusion criteria. Put simply, we encourage researchers to report this information at the level of detail that would be sufficient for an independent researcher to potentially replicate the dataset construction process.

Another concern regarding archival data transparency revolves around accessibility. Archival data can be open and publicly available or private and proprietary. When archival data is open and publicly available, there is little concern about data transparency. In fact, publicly accessible data facilitate attempts to replicate research findings with the same sample and procedures. Having many analysts is beneficial as they often have different approaches to analyzing datasets, which can have surprisingly drastic effects on the inferences drawn (Silberzahn & Uhlmann, 2015). With a group of analysts and a thorough discussion, it is less likely that a paper will pursue an outlier approach that may lead the field astray. This is possible with any dataset, but a publicly available archival dataset enables this across different groups of researchers.

Conversely, significant concerns arise when archival data are private and proprietary. Data sharing is necessary in order to enable verifying the correctness and reproducibility of data analyses, as well as for investigation purposes in cases where fraud is suspected. Although we acknowledge the risks associated with data sharing, in particular the loss of confidentiality (Zimmer, 2010), we strongly believe that there are many approaches that researchers can take to overcome these risks and challenges. There is an emerging consensus that researchers should only be required to make available a “minimal dataset” that includes “data that are relevant to the specific analysis presented in the paper” (Silva, 2014), instead of the entire dataset. We thus highly encourage researchers to negotiate terms with their proprietary data providers that would allow for greater data openness. To alleviate concerns about the confidentiality of the organization and its members, researchers can clarify the extent of data sharing by clearly delineating the steps taken to de-identify the data (see Mackinnon, 2014 for a guide on data de-identification). Researchers can also propose that they would upload the dataset onto a secure data repository but only provide access to other researchers with the agreement that they do not disseminate the data. Given the reduced risks related to data confidentiality that should accompany such restricted access, the hope is that data providers and social psychology researchers would be comfortable with sharing their datasets. In sum, although we may not be able to achieve “perfect” pre-registration, openness, and transparency with pre-existing data, we can strive toward achieving this by being as transparent as possible, which would at the very least make salient the potential biases of the researcher.

2.1.2. Longitudinal data

Using an archival approach further facilitates the examination of phenomena over time. As noted by Barnes, Dang, Leavitt, Guarana, and Uhlmann (2018), archival research is a potentially useful tool to investigate the issues associated with time in social psychology research. As many archival datasets have longitudinal designs, researchers are able to study the effect of durations and the trajectory of phenomena. Temporal factors may influence either the standings of variables or the relationships among different variables. For instance, personality research has consistently demonstrated that personality change occurs

across one's lifespan (Helson, Jones, & Kwan, 2002; Roberts, Walton, & Viechtbauer, 2006). However, these temporal changes in personality are not accounted for in cross-sectional data. Given that cross-sectional data is frequently collected in social psychological research, opportunities to examine phenomena over time is limited in the field of social psychology.

The nature of many archival datasets lends itself to longitudinal studies. In a skill acquisition study, researchers analyzed the training history and performance of > 800,000 online game players over two months (Stafford & Dewar, 2014). They found a relationship between practice amount and spacing with subsequent performance, thus confirming prior experimental findings on skill acquisition. In another example, Back, Kűfner, and Egloff (2010) conducted a study examining the emotional timeline of the September 11 attacks. Back and colleagues analyzed the use of emotional words in text messages (sadness, anxiety, and anger) over the course of the day, taking into account the specific events that occurred (e.g., plane crashes, revealing of suspect, announcement of killed firefighters, etc.). Across the terrorist attack event, the researchers found a dynamic pattern of negative emotions in response to the traumatic event: people initially did not react with sadness, experienced several anxiety outbursts but recovered quickly, and steadily became angrier. As these examples illustrate, researchers can leverage archival data to examine social behaviors over an extended period of time.

2.2. Sample Characteristics

2.2.1. Increased diversity of samples

Another benefit of using an archival research approach is the ability to obtain samples from a much broader variety of sources as compared to those typically available for laboratory research. The use of a variety of samples increases confidence in the generalizability of the research findings to the larger population. Whereas laboratory samples tend to consist of undergraduate psychology students or other forms of paid subject pools such as Mechanical Turk, archival research is able to capture more diverse samples from within the world population. For example, in the Sleep and Cyberloafing Study, the researchers obtained the relative percentage of internet searches in the Entertainment category across the U.S. metropolitan population. These data were retrieved from the Google archival database and were therefore much more diverse and representative than a relatively homogeneous university subject pool or opt-in research service, or even than a single field study. Similarly, in the Facebook A/B Study, all people who viewed Facebook in English qualified for selection into the experiment. This means that the study sample is likely to be much more diverse in terms of demographics, such as country of origin, ethnicity, age, and occupation, as compared to laboratory experiments and field surveys.

Researchers have also used a variety of archival datasets in their research, such as data from professional sports players (Swaab, Schaefer, Anicich, Ronay, & Galinsky, 2014), online game players (Stafford & Dewar, 2014), and expert mountain climbers (Anicich, Swaab, & Galinsky, 2015). Although archival research is advantageous in its potential to access diverse populations, it is important to note the misconception that sampling biases would be less systematic or problematic. Sampling biases are still dependent on the individuals responsible for the data collection and do not cease to exist when the data are used in an archival manner. These biases may also manifest in different forms. For example, in the Anticipatory Consumption Study, biases may have arisen in the form of people representing themselves differently in media interviews than they would naturally behave, or reporters curating their articles in an inaccurate manner because they have the intention to create sensational news reports.

2.2.2. Employment of large datasets

Although not always the case, archival research has the potential to involve large sample sizes that often include thousands (or even

millions) of participants. This stands in stark contrast to laboratory experiments which often include fewer than a hundred participants. Because large datasets increase statistical power and ensure that estimates are more precise, archival research is useful in mitigating the concerns that are common to small samples. For example, the Facebook A/B Study had a sample of > 3 million Facebook posts. Other large archival datasets include social media and government data. Using a sample of 3.8 million Twitter users and a dataset of close to 150 million tweets, Barberá, Jost, Nagler, Tucker, and Bonneau (2015) examined how peoples' ideological preferences influenced discussions about a range of political and non-political issues. However, a drawback to the statistical power afforded by these large archival datasets is that very small effects will reach statistical significance but may not be practically significant. It is thus important to go beyond significant *p*-values to also consider the meaningfulness of effect sizes. In the case of the Facebook A/B Study, the findings were as statistically significant as $p < .0001$, but the effect sizes from the manipulation were as small as $d = 0.001$. This has aroused concerns about the meaningfulness of the finding that emotions can spread throughout an online social network.

On the other hand, depending on the dataset and research question at hand, small effect sizes can be meaningful. Some datasets have implications for largescale estimates of the effect of the phenomenon of study. For instance, it has been estimated that modest (approximately 5%) increases in acute myocardial infarction (Janszky & Ljung, 2008), cyberloafing (Wagner et al., 2012), and workplace injuries (Barnes & Wagner, 2009) as a result of the switch to Daylight Saving Time cost the US economy \$434 million every year (Chmura, 2016). In these cases, the findings have practical significance as even a small increase in heart attacks and injuries aggregates to a huge cost in terms of employee safety, well-being, and productivity.

2.3. Type of measures

2.3.1. Realism

With an archival research approach, social psychologists can obtain behavioral data from real-world contexts, such as sports games, online games, politics, and social media posts. The Facebook A/B Study used a sample of actual Facebook posts and the Sleep and Cyberloafing Study examined actual Google searches. In another study utilizing data from > 50,000 Major League Baseball games, Larrick, Timmerman, Carton, and Abrevaya (2011) demonstrated how high temperatures and provocation, measured by how often the pitcher's teammates had been hit by the opposing team earlier in the game, increased retaliation in a sports game setting, measured by the likelihood that batters were hit by a pitch. Likewise, researchers have analyzed archival datasets of mountain climbers, National Basketball Association players, and military cadets, in the attempt to understand the determinants of individual and group performance (Anicich et al., 2015; Swaab et al., 2014; Wrzesniewski et al., 2014). Notably, the performance data obtained from archival datasets are from real-world behaviors as compared to laboratory experiments where participants are instructed to engage in tasks that are necessarily constrained to the context of the laboratory (Bélanger, Lafreniere, Vallerand, & Kruglanski, 2013; Tenney, Logg, & Moore, 2015), and may lack complete fidelity to real-world contexts. As such, a benefit of an archival approach is realism, which increases external validity by ensuring that predicted relationships hold true in real-life circumstances.

2.3.2. Reduced demand characteristics

Archival data collections often occur surreptitiously, which reduces demand characteristics. Demand characteristics are experimental artifacts that arise when participants attempt to discern the experimenter's hypotheses and consciously or unconsciously change their behavior in response to their interpretation of the research purpose (Orae, 1969). For example, participants may play the "good-participant role" by behaving in a manner that would confirm what they presume is the

research hypotheses. Participants may also play the "bad-participant role" by deliberately behaving in a manner that contradicts their guess of the hypothesis. Demand characteristics thus result in participants behaving in a way that differs from how they would behave in a natural scenario. This difference may be exacerbated by experimenter biases, such as differential treatment of participants, as participants may make inferences about the research purpose by observing the experimenter's behavior (Klein et al., 2012).

Archival research circumvents these problems because participants are often either unaware of the ongoing data collection, or of the purposes for which the data are collected. In such cases, there are often minimal (sometimes even zero) demand characteristics. In the Facebook A/B Study, participants were unaware that their Facebook feed was manipulated, and in the Sleep and Cyberloafing Study, participants were unaware that their Google searches were examined. Researchers have also utilized other surreptitiously collected archival data such as eBay bid histories (Ku, Galinsky, & Murnighan, 2006) and tweets (Barberá et al., 2015). In the Anticipatory Consumption Study where researchers examined news story archives about people waiting in line, these individuals may have been aware that they were being observed, but the observation was not seen as being driven by a research question. In these cases where research is conducted using an archival dataset, experimenters have no interaction and thus no avenue to influence the participants. At the same time, it is less salient to participants that they are being observed or that their responses and behaviors are being utilized for research purposes. As such, the use of archival data often avoids the introduction of demand characteristics that would otherwise have been present in laboratory and experience sampling studies. Nonetheless, as we noted earlier, it is important to acknowledge that other biases may be introduced in the archival research data collection process. For instance, in the Anticipatory Consumption Study, reporter prejudices and peoples' desire to present their best selves to the media could introduce biases in the archival data.

Researchers should also be thoughtful about the ethical concerns associated with covert data collection. For example, the mood manipulation in the Facebook A/B Study sparked controversy because it raised questions about informed consent and user privacy. In an editorial expression of concern with respect to the Facebook A/B Study, Verma (2014, p. 10779) appropriately noted that this emerging area of social media archival research "needs to be approached with sensitivity and with vigilance regarding personal privacy issues." On the one hand, Facebook users have agreed to their Data Use Policy prior to creating a Facebook account, and this agreement may constitute informed consent for this research. However, the failure to overtly obtain consent and provide participants with the opportunity to opt out of the study violates federal policy for the protection of human subjects (i.e., the "Common Rule", United States Department of Health and Human Services, n.d.). Similarly, public outcry around the Cambridge Analytica saga stemmed from the lack of an overt consent process for a "psychographics research study" on Amazon's Mechanical Turk, which led to the covert collection and misuse of the Facebook data of > 30 million Facebook users (Bump, 2018). These examples highlight the need for researchers to be especially cautious when taking advantage of the surreptitious nature of archival research.

2.3.3. Ability to ethically study socially sensitive phenomena

A significant strength of archival research over other methodological designs is the ability to examine socially sensitive phenomena in an ethical manner. It is often difficult and unethical to conduct laboratory experiments or field research on certain socially sensitive topics, including illegal behavior, highly personal behavior, violence, vice, and death. Archival research can sometimes circumvent this problem by measuring socially sensitive behaviors in a non-confrontational and indirect manner.

For example, a study by Leader, Mullen, and Abrams (2007) used archival data in the form of newspaper reports and photographs to

examine lynching—an illegal and highly violent activity in which White mobs murdered African-Americans in the United States. They found, among other things, that lynch mob atrocity was intensified by the mob size and was not influenced by the relative number of African Americans in the community. In an interesting study, Calogero and Mullen (2008) examined the facial prominence of George W. Bush in political cartoons across two different wars. They found that Bush had lower facial prominence after the onset of the war, suggesting that he was portrayed to be less powerful and dominant. The studies mentioned could only be conducted in an archival manner and not in the laboratory because it would be unethical to manipulate participants to engage in lynching or to start wars.

2.3.4. Data validity and reliability concerns

A drawback of capturing actual behavior in the real world is the difficulty interpreting many kinds of indirect measures. As archival datasets are mostly compiled for reasons other than academic research, it is generally more difficult to assess their psychometric properties as compared to data compiled by other research methods. For example, academic scores are collated for college admissions and sports data are collected to facilitate player evaluations and to inform training and game strategies. In some cases, only single-item measures are collected, making it problematic to statistically assess reliability and validity. There is also a huge body of externally valid archival data that researchers can capitalize on, such as television shows, newspaper archives, photographs, artwork, diaries, and letters. Whereas some archival datasets have such extensive construct validity problems that they are not useful for research, other archival datasets have few or no construct validity problems. Most archival datasets likely fall between these two extremes, and researchers should thus be cautious in evaluating these datasets.

Indeed, the issues raised are especially problematic for psychologists because studying human cognition and psychological processes entail remote inferences about psychological states from these real-world data. For example, in the Sleep and Cyberloafing study, the authors argue that people who search for words in the Entertainment category are cyberloafing, but it is unclear what mechanism connects lost sleep and cyberloafing—which could arguably be self-regulatory depletion, a lack of motivation, or negative affect. However, social psychologists can more confidently examine psychological states by ensuring that their variables are clearly operationalized. For instance, the Facebook A/B study researchers and Back et al. (2010) gathered data on expressed emotions by analyzing social media posts and text messages. Comparatively, this approach is more rigorous than that in the Sleep and Cyberloafing study as it uses well-validated tools that were specially developed for sentiment analysis. This suggests that researchers should critically consider whether the psychological states in which they are interested in can be accurately captured in archival research.

Researchers can also use statistical methods to increase the strength of their arguments by (1) using analyses to rule out confounding variables and alternative explanations and (2) conducting supplementary laboratory studies to verify inferred psychological states and processes. In the Sleep and Cyberloafing study, the researchers controlled for the linear trend in cyberloafing behaviors from February through April to account for people being more likely to spend time outdoors because of warmer spring weather. In the Anticipatory Consumption Study where researchers coded and analyzed news stories to examine the relationship between type of purchase and people's mood or behavior while waiting in line, they ran further analyses to rule out the possible explanation of scarcity of the item influencing people's moods and behaviors. They further conducted an experiment to strengthen their evidence for causality by randomly assigning participants to recall a specific instance when they waited in a long line to purchase either a material good or an experiential good. Then, participants reported how pleasant the experience of waiting had been. These strategies may not completely solve data validity and reliability concerns, but are certainly helpful for researchers in strengthening their evidence for causality and confidence in their inferences.

3. Starting points, resources, and directions for archival research

In this section, we provide some direction for psychologists who may be interested in delving into archival research, but are unsure of where to begin. We point readers to resources to get started, introduce some useful analytical techniques that will be helpful to deal with archival datasets, and finally, share some recent novel and interesting archival research approaches that have been observed in the field. Our intention is not to provide a comprehensive manual for archival research, but rather, to expose readers to the possibilities of archival research and how they can set about utilizing it.

3.1. Data acquisition and processing

Working with archival data requires that researchers have the skills to successfully acquire, process, and analyze the data. There are plenty of archival data sources that are often free and publicly accessible. In searching for a suitable archival dataset, we strongly urge researchers to be creative in producing innovative research designs. This can be facilitated by brainstorming about how constructs of interest may be naturally manipulated or measured in the real world. Based on the phenomena that the researcher is interested in, we encourage the exploration of various archival data sources, including data from social media, sports teams, population census, books, and newspaper articles. In cases in which researchers are constructing their archival dataset, one needs to be able to efficiently extract, catalogue, and organize large datasets from websites. Finally, to process the data, researchers need to

Table 2
 A summary of available resources for archival research.

Archival research starting points	Examples
Free and publicly available data sources	<ul style="list-style-type: none"> • Social media data (e.g., Twitter; Facebook; Instagram) • Sports data (e.g., National Football League; Major League Baseball; National Basketball Association) • Communication and media records (e.g., Newspaper archives; video recordings; press releases) • Government databases (e.g., U.S. Census; Bureau of Labor Statistics) • eBay auction data
Data retrieval tools	<ul style="list-style-type: none"> • Google Trends provides data on Google search term frequency • Web scraping (common coding languages include R, Python, and Java; useful to learn Wget and cURL to simulate navigation of human-readable websites via web browser) • Application programming interfaces connect softwares (e.g., R and Python) with company databases (APIs; e.g., WikipediR, TwitteR, instaR and Rfacebook) • Programming languages that facilitate data-retrieval from entire databases (e.g., SQL, MySQL, PostgreSQL)
Data processing and data visualization tools	<ul style="list-style-type: none"> • Tableau • Microsoft Power BI • Multiple R packages (e.g., reshape, reshape2, ggplot2, dplyr)

effectively sort through, index, analyze, and present large datasets that are potentially stored across multiple devices. In Table 2, we provide a non-exhaustive list of resources available to researchers as they acquire, manage, and utilize archival data. A helpful recent paper by Braun, Kuljanin, and DeShon (2017) also provides an in-depth discussion on how researchers can acquire and process big datasets.

3.2. Useful analytical techniques

The determination of cause and effect requires covariation, temporal precedence, and the elimination of plausible alternative explanations (Rosenthal & Rosnow, 1991). These requirements apply to research using archival datasets just as they do to experimental studies. Of the four examples we highlight in this paper, causality can only be unambiguously asserted for the Facebook A/B Study (true experiment), and evidence for causality is stronger for the Divorce Education Program Study (quasi-experiment) and the Sleep and Cyberloafing Study (natural experiment) as compared to the Anticipatory Consumption Study (correlational study). The field's limited use and familiarity with archival research could be related to a lack of sophistication in our understanding of archival research and datasets, leading to erroneous inferences of causality. Hence, we recommend that researchers employ complementary analytical techniques to strengthen evidence for causality. Here, we introduce three useful techniques that can help to strengthen evidence for causality.

3.2.1. Autoregressive models

In autoregressive models, earlier values of the dependent variable in a time series, also known as lagged variables, are used as independent variables in the regression model. This technique helps to account for lagged effects, wherein what happened in the past influences the future. For example, in their study of the effect of increasing national income on subjective well-being, Diener, Tay, and Oishi (2013) used this technique to account for autocorrelations among country waves that tend to be present in time series data (Raudenbush & Bryk, 2002), making it a more statistically optimal approach compared to traditional ordinary least squares regression. By considering temporal dynamics, this technique allows for stronger evidence of causality through the separation of the dynamic portion of the model (i.e., the relationships between the time-lagged values of the variables) from the simultaneous portion (i.e., the contemporaneous values), which assists in inferences about the temporal order of the effects (Rosmalen, Wenting, Roest, de Jonge, & Bos, 2012; see Brandt & Williams, 2007, for an in-depth discussion).

3.2.2. Propensity score matching and use of control groups

When using observational data, it may be unclear whether observed changes in the outcome are attributed to the treatment or to an uncontrolled confounding variable. Propensity score matching removes selection effects in the composition of treatment and control groups by creating a control group that is similar to the treatment group based on the pre-treatment characteristics of participants. For example, in examining whether important romantic relationship transitions account for why individuals differ in their self-esteem trajectories, Luciano and Orth (2017) matched participants who experienced a transition (treatment group) to participants who did not experience a transition (control group) using calculated propensity scores. These propensity scores reflect the likelihood of an individual experiencing a future event based on all the scores that the individual has on potential confounding variables (e.g., personality and several demographic variables). By allowing researchers to control for a large set of confounding variables, this technique increases the ability to attain a causal interpretation of effects because any differences in the outcome can more likely be attributed to the treatment (see Salkind, 2010, for an in-depth discussion).

3.2.3. Longitudinal hierarchical linear models with time-varying covariates

It is theoretically impossible to be sure that researchers have accounted for all possible confounding variables in their analysis. A partial solution to this problem is the longitudinal growth-curve modeling technique used with hierarchical linear models, wherein each subject serves as his or her own control, thereby eliminating between-individual confounding variables. For example, Duckworth, Tsukayama, and May (2010) used this technique to show that within-individual changes in self-control over time predicted subsequent changes in students' grades, but not vice versa. However, in spite of the usefulness of this technique, it is important to note that this analytic method only effectively controls for time-invariant confounds and cannot rule out time-varying confounds that change in sync with the studied variables. Preacher, Wichman, MacCallum, and Briggs (2008) provide a detailed discussion of this technique.

3.3. Novel and interesting archival research approaches

Thus far, we are appreciative of the various types of archival research that have been used in social psychology, including the use of data from social media, sports teams, population censuses, and other large-scale government agency data collection efforts. In particular, there are some archival research endeavors that have stood out for their novelty and creativity. For example, researchers have used natural language processing to extract psychological information of playwrights from their plays (Boyd & Pennebaker, 2015) and analyzed college admissions essays to predict academic success (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014).

Other innovative approaches to archival research include the gathering of online game data, such as *blitz chess*, *Balderdash*, and *Axon*, which allows researchers to examine skill acquisition and performance (Alter & Oppenheimer, 2008; Burns, 2004; Stafford & Dewar, 2014). We challenge social psychologists to be creative about how they study their phenomenon of interest. For example, in examining the importance of moral character and warmth in person perception and evaluation, Goodwin, Piazza, and Rozin (2014) complemented traditional research methods with an archival dataset of obituaries. Also, Langner and Winter (2001) examined government-to-government archival documents from various crises to study the motivational basis of compromise making. These are excellent examples of thinking outside the box when it comes to utilizing an archival research approach, and we enthusiastically encourage other social psychologists to follow suit.

3.4. Capitalizing on the potential of interdisciplinary collaborations

We acknowledge that the effective use of archival research sometimes requires the mastery of difficult technical skills such as web scraping and machine learning. These technical requirements may dissuade some social psychologists from delving into archival research. Another viable option would be to enter interdisciplinary collaborations with colleagues from other disciplines, such as computer science and engineering, who have the expertise to work with big data effectively. There are many benefits to interdisciplinary collaborations. Interdisciplinary research encourages researchers to tackle a problem from diverse and unique angles. This encourages creativity and novel perspectives. Interdisciplinary collaborations could also increase researchers' chances of attaining funding. Many research grant agencies, including the National Science Foundation (NSF) and National Institutes of Health (NIH), strongly advocate for interdisciplinary research and have specific funding programs targeting such interdisciplinary research projects (NSF, n.d.; NIH, 2017). Therefore, social psychologists who are interested in adopting an archival research approach, but are less confident or keen on mastering challenging data processing and management techniques, should consider capitalizing the many benefits of interdisciplinary research.

4. Conclusion

There are two basic purposes of research design: first, to provide answers to research questions, and second, to control variance (Kerlinger, 1986). Although laboratory experiments are key to achieving the latter, utilizing laboratory experiments as the only research paradigm may obscure the accuracy of the former due to its inevitable limitations. In this paper, we advocate that as social psychologists we should increase our usage of archival research while still retaining laboratory experiments and field studies as extremely valuable research paradigms. We do not advocate that every paper in social psychology use archival research designs. This is an unrealistic expectation. A more reasonable expectation is that across a given program of research, social psychology researchers would include a healthy mix of laboratory, field, and archival research. Convergence in findings across the use of various research methodologies provides stronger support for the hypothesized relationships as compared to the use of only one research method.

With the recent data explosion and vast technological advancements, archival research as a methodological approach has never been more promising. In this paper, we have provided a balanced discussion of the pros and the cons of archival research. We highlight the benefits of archival research that are not achievable with traditional research methods, such as ethically studying socially sensitive phenomena, and the utilization of larger and more diverse samples, but also acknowledge its weaknesses, including pre-existing biases from the initial collection and construction of the archival dataset, other forms of biases, and data validity and reliability concerns. Taken together, we strongly urge social psychologists to capitalize on the many benefits of archival research—a research approach that has been underappreciated in the field of social psychology. By adding archival research to our methodological toolkit, social psychologists can utilize archival research together with traditional research methods, thereby increasing the rigor of our research and fortifying the body of knowledge in the field.

References

Alter, A. L., & Oppenheimer, D. M. (2008). Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). *Psychological Science*, 19, 161–167.

Anicich, E. M., Swaab, R. I., & Galinsky, A. D. (2015). Hierarchical cultural values predict success and mortality in high-stakes teams. *Proceedings of the National Academy of Sciences*, 112, 1338–1343.

Back, M. D., Küfner, A. C., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, 21, 1417–1419.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26, 1531–1542.

Barnes, C. M., Dang, C., Leavitt, K., Guarana, C., & Uhlmann, E. L. (2018). Archival data in micro organizational research: A toolkit for moving to a broader set of topics. *Journal of Management*, 44, 1453–1478.

Barnes, C. M., & Wagner, D. T. (2009). Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of Applied Psychology*, 94, 1305–1317.

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153–158.

Bélanger, J. J., Lafreniere, M. A. K., Vallerand, R. J., & Kruglanski, A. W. (2013). Driven by fear: The effect of success and failure information on passionate individuals' performance. *Journal of Personality and Social Psychology*, 104, 180–195.

Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science*, 26, 570–582.

Brandt, P. T., & Williams, J. T. (2007). *Multiple time series models*. Thousand Oaks, CA: Sage Publications.

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2017). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods* (1094428117690235).

Bump, P. (2018). *Everything you need to know about the Cambridge Analytica-Facebook debacle*. The Washington Post. Retrieved from: https://www.washingtonpost.com/news/politics/wp/2018/03/19/everything-you-need-to-know-about-the-cambridge-analytica-facebook-debacle/?utm_term=.5f41cf375082.

Burns, B. D. (2004). The effects of speed on skilled chess performance. *Psychological Science*, 15, 442–447.

Calogero, R. M., & Mullen, B. (2008). About face: Facial prominence of George W. Bush in

political cartoons as a function of war. *The Leadership Quarterly*, 19, 107–116.

Cherlin, A. (1991). On analyzing other people's data. *Developmental Psychology*, 27, 946–948.

Chmura, C. (2016). March 11. *SleepBetter lost-hour economic index*. Retrieved from: <http://www.chmuraecon.com/blog/2016/march/11/sleepbetter-lost-hour-economic-index/>.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15, 687–693.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimental design and analysis issues for field settings*. Chicago: Rand McNally.

deLusé, S. R., & Braver, S. L. (2015). A rigorous quasi-experimental design to evaluate the causal effect of a mandatory divorce education program. *Family Court Review*, 53, 66–78.

Diener, E., Tay, L., & Oishi, S. (2013). Rising income and the subjective well-being of nations. *Journal of Personality and Social Psychology*, 104, 267–276.

Doss, B. D., Rhoades, G. K., Stanley, S. M., & Markman, H. J. (2009). The effect of the transition to parenthood on relationship quality: An 8-year prospective study. *Journal of Personality and Social Psychology*, 96, 601–619.

Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychological and Personality Science*, 1, 311–317.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326, 535–538.

Fraley, C., & Vazire, S. (2014). The n-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS One*, 9(10), e109019.

Gallo, A. (2017). A refresher to A/B testing. Retrieved from: <https://hbr.org/2017/06/a-refresher-on-ab-testing>.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106, 148–168.

Hales, A. H. (2016). Does the conclusion follow from the evidence? Recommendations for improving research. *Journal of Experimental Social Psychology*, 66, 39–46.

Helson, R., Jones, C., & Kwan, V. S. (2002). Personality change over 40 years of adulthood: Hierarchical linear modeling analyses of two longitudinal samples. *Journal of Personality and Social Psychology*, 83, 752–766.

Hollenbeck, J. R., DeRue, D. S., & Mannor, M. (2006). Statistical power and parameter stability when subjects are few and tests are many: Comment on Peterson, Smith, Martorana, and Owens (2003). *Journal of Applied Psychology*, 91, 1–5.

International Data Corporation (2014). The digital universe of opportunities: Rich data and the increasing value of the internet of things. Retrieved from <https://ireland.emc.com/leadership/digital-universe/2014view/index.htm>.

Janszky, I., & Ljung, R. (2008). Shifts to and from daylight saving time and incidence of myocardial infarction. *New England Journal of Medicine*, 359, 1966–1969.

Kerlinger, F. (1986). *Foundations of behavioral research*. Fort Worth, TX: Harcourt.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196–217.

Klein, O., Doyen, S., Leys, C., de Gama, P. A. M. D., Miller, S., Questienne, L., & Cleeremans, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspectives on Psychological Science*, 7, 572–584.

Kramer, A. D., 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, 8788–8790.

Ku, G., Galinsky, A. D., & Murnighan, J. K. (2006). Starting low but ending high: A reversal of the anchoring effect in auctions. *Journal of Personality and Social Psychology*, 90, 975–986.

Kumar, A., Killingsworth, M. A., & Gilovich, T. (2014). Waiting for merlot: Anticipatory consumption of experiential and material purchases. *Psychological Science*, 25, 1924–1931.

Langner, C., & Winter, D. G. (2001). The motivational basis of compromise and concessions: Archival and experimental studies. *Journal of Personality and Social Psychology*, 81, 711–727.

Larrick, R. P., Timmerman, T. A., Carton, A. M., & Abrevaya, J. (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science*, 22, 423–428.

Leader, T., Mullen, B., & Abrams, D. (2007). Without mercy: The immediate impact of group size on lynch mob atrocity. *Personality and Social Psychology Bulletin*, 33, 1340–1352.

Luciano, E. C., & Orth, U. (2017). Transitions in romantic relationships and development of self-esteem. *Journal of Personality and Social Psychology*, 112, 307–328.

Mackinnon, S. *Privacy in the age of open data*. (2014). Retrieved from <http://osc.centerforopenessence.org/2014/01/29/privacy-and-open-data/> (January 29).

MIT Technology Review (2013). Big data gets personal. Retrieved from: <http://www.datascienceassn.org/sites/default/files/Big%20Data%20Gets%20Personal.pdf>.

NIH. *Interdisciplinary*. (2017). Retrieved from <https://commonfund.nih.gov/Interdisciplinary> (May 5).

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Contestabile, M. (2015). Promoting an open research culture. *Science*, 348, 1422–1425.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 201708274.

NSF. How does NSF support interdisciplinary research?. Retrieved from [https://www.nsf.gov/od/oa/additional_resources/interdisciplinary_research/support.jsp\(n.d.\)](https://www.nsf.gov/od/oa/additional_resources/interdisciplinary_research/support.jsp(n.d.)).

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349.

Orae, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal, & R. L. Rosnow (Eds.). *Artifact in behavioral research* (pp. 143–179). New

- York: Academic Press.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS One*, 9, e115844.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Quantitative applications in the social sciences: Latent growth curve modeling*. Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Remler, D. K., & Van Ryzin, G. G. (2015). *Research methods in practice: Strategies for description and causation*. London: Sage.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132, 1–25.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. Boston: McGraw Hill.
- Rosmalen, J. G., Wenting, A. M., Roest, A. M., de Jonge, P., & Bos, E. H. (2012). Revealing causal heterogeneity using time series analysis of ambulatory assessments: Application to the association between depression and physical activity after myocardial infarction. *Psychosomatic Medicine*, 74, 377–386.
- Sales, S. M. (1973). Threat as a factor in authoritarianism: An analysis of archival data. *Journal of Personality and Social Psychology*, 28, 44–57.
- Salkind, N. J. (2010). *Encyclopedia of research design*. Thousand Oaks, CA: Sage Publications.
- Sbarra, D. A., Emery, R. E., Beam, C. R., & Ocker, B. L. (2014). Marital dissolution and major depression in midlife: A propensity score analysis. *Clinical Psychological Science*, 2, 249–257.
- Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43, 1248–1264.
- Shultz, K. S., Hoffman, C. C., & Reiter-Palmon, R. (2005). Using archival data for IO research: Advantages, pitfalls, sources, and examples. *The Industrial-Organizational Psychologist*, 42, 31–37.
- Silberzahn, R., & Uhlmann, E. L. (2015). Many hands make tight work: Crowdsourcing research can balance discussions, validate findings and better inform policy. *Nature*, 526, 189–191.
- Silva, L. *PLOS' new data policy: Public access to data*. (2014). Retrieved from <http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/> (February 24).
- Stafford, T., & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, 25, 511–518.
- Swaab, R. I., Schaefer, M., Anicich, E. M., Ronay, R., & Galinsky, A. D. (2014). The too-much-talent effect team interdependence determines when more talent is too much or not enough. *Psychological Science*, 25, 1581–1591.
- Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology*, 108, 377–399.
- United States Department of Health and Human Services (n.d.). Federal policy for the protection of human subjects ('common rule'). Retrieved from: <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>.
- van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2–12.
- Verma, I. M. (2014). Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 10779.
- Wagner, D. T., Barnes, C. M., Lim, V., & Ferris, D. L. (2012). Lost sleep and cyberloafing: Evidence from the laboratory and a daylight saving time quasi-experiment. *Journal of Applied Psychology*, 97, 1068–1076.
- Wrzesniewski, A., Schwartz, B., Cong, X., Kane, M., Omar, A., & Kolditz, T. (2014). Multiple types of motives don't multiply the motivation of west point cadets. *Proceedings of the National Academy of Sciences*, 111, 10990–10995.
- Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12, 313–325.