facebook as a research tool

A look at how to recruit participants using Facebook — and the ethical concerns that come with social media research.

By Michal Kosinski, PhD, Sandra C. Matz, Samuel D. Gosling, PhD, Vesselin Popov and David Stillwell, PhD
Overview

CE credits: 1  
Exam items: 10  
Learning objectives: After completing this course participants will be able to:  
1. Understand the opportunities and challenges of Facebook research.  
2. Learn practical ways to conduct research on Facebook.  
3. Understand the ethical considerations of such research.  

Facebook has become a significant part of daily life for nearly 1.4 billion people worldwide. While many researchers have explored Facebook’s influence on individuals and societies, its potential as a powerful research tool has been largely overlooked.  

Access to the large and diverse samples offered by Facebook could help to address a major challenge in social science: its overreliance on samples that are relatively small, student and WEIRD (Western, educated, industrialized, rich and democratic). Furthermore, Facebook can be used to circumvent the limitations of self-reports and laboratory-based studies by providing access to records of actual behavior expressed in a natural environment.  

In this article, we review the opportunities and challenges of Facebook research; provide a number of practical recommendations for effectively conducting research within this environment; and discuss several ethical considerations. We hope to convince the reader that using Facebook in research generally produces robust results and can be as easy as posting an advertisement on Facebook or adding a “Log in with Facebook” button to an online survey.

Part #1: Recruiting participants

While the Facebook population is not perfectly representative since its users tend to be younger and better educated than the general population, its sheer size implies that even underrepresented groups are relatively large. For example, as of 2014, nearly 35 percent of Americans over 65 years of age were on Facebook, and their number was quickly growing.

One of the least expensive and most efficient ways to dip into Facebook’s participant pool is by snowball sampling: convincing Facebook users to recruit their friends to join a study. If enough participants do so, the positive feedback loop may lead to self-sustaining studies with a rapid growth in sample size. In order to go viral, a study must be engaging to its participants (such as by including a game or offering compelling feedback) and must integrate the invitation of friends as a core part of the experience. A recent myPersonality study, for example, offered a 360-degree assessment feature, encouraging participants to invite friends to judge their personality. This study, originally shared with the author’s 150 Facebook friends, went viral and attracted over six million participants in four years.

Advantages of snowball sampling include low costs and large sample sizes. One of the downsides is that the first participants are likely to disproportionately affect the composition of the sample, since people tend to interact with those similar to themselves. Furthermore, people with many friends are more likely to be recruited into the sample. (The size and diversity of the Facebook population can certainly help to minimize this disadvantage; given enough participants, the representativeness of the population can be improved by weighting.)

An effective alternative to snowball sampling is offered by targeted advertising. The Facebook advertising platform can be used to target audiences defined by a wide range of preferences (such as liking “getting up early in the morning”); behaviors (liking “running”); and demographic variables including location, education, language, political views, ethnicity, sexual orientation, income and many more. This approach can be used to obtain representative samples or reach “rare” participants, such as those stigmatized in the offline world or those who are hesitant to meet researchers face-to-face.

Research shows that Facebook advertising reduces the costs of targeted participant recruitment. Not only did Facebook ads outperform traditional methods, such as postal surveys, but they are also more cost-efficient than Google advertising, online newsletters and emails. In 10 recent studies using Facebook advertising, the average cost per participant was $13.75. Furthermore, Facebook can be used to remain in contact with ex-participants. For example, over 150,000 people subscribed to the myPersonality project’s Facebook page over the years. Their comments provided the authors with invaluable feedback on the design of and issues with the studies. Also, messages published on the project’s page attracted considerable attention — in a few hours, it was possible to recruit tens of

Welcome to ‘CE Corner’

“CE Corner” is a continuing education article offered by the APA Office of CE in Psychology. This feature will provide you with updates on critical developments in psychology, drawn from peer-reviewed literature and written by leading psychology experts. To earn CE credit, after you read this article, purchase the online exam at www.apa.org/education/ce/1360399.aspx.

Upon successful completion of the test — a score of 75 percent or higher — you can immediately print your CE certificate. The test fee is $25 for members and $35 for nonmembers. The APA Office of CE in Psychology retains responsibility for the program. For more information, call (800) 374-2721.
Research shows that Facebook advertising reduces the costs of targeted participant recruitment. Not only did Facebook ads outperform traditional methods, such as postal surveys, but they are also more cost-efficient than Google advertising, online newsletters and emails.

and such profiles are relatively easy to detect. For example, real users accumulate their friends and likes over time, whereas fake profiles are likely to be populated with likes and friends in a single burst of activity.

There is a solid and growing body of empirical evidence supporting the validity of information contained in Facebook profiles. However, as some parts of the profile are self-reported and others can be selectively removed by the profile owner, these data are not entirely free of biases, such as social desirability and intentional misrepresentation. The Facebook user experience is also highly personalized by the algorithms selecting news feed stories, targeting advertisements, and recommending new friend connections. As users are more likely to interact with content suggested to them by Facebook, the extent to which these algorithms affect users’ behavior cannot be known fully.

Furthermore, profile entries such as job position or political views are recorded using open-text input fields, which can be equipped with autocomplete features that suggest entries based on the first few letters typed in. This may minimize spelling errors, but it can also introduce biases. For example, a user might have intended to identify himself as a “social psychologist,” but may settle on “social scientist” if Facebook suggests the latter. Finally, Facebook’s constant evolution as a platform also drives changes in users’ behavior. Status updates, for instance, were originally preceded with the user's name (i.e., “Alice Miller is...”), encouraging updates written in the third person. Nowadays, status updates are free to contain any text.

Part #2: Recording Facebook profile data

Facebook profile information includes self-reported information (such as schools attended, current workplace, age and gender); traces of behavior (such as status updates or likes); and data contributed by others (such as photo tags or comments on a user’s wall). These data can be recorded retrospectively and thus help researchers to address the shortcomings of participants’ memories and biases. Also, while some researchers are concerned by the risk of participants using fake Facebook profiles to join the study, this is rarely the case because real users accumulate their friends and likes over time, whereas fake profiles are likely to be populated with likes and friends in a single burst of activity.

There is a solid and growing body of empirical evidence supporting the validity of information contained in Facebook profiles. However, as some parts of the profile are self-reported and others can be selectively removed by the profile owner, these data are not entirely free of biases, such as social desirability and intentional misrepresentation. The Facebook user experience is also highly personalized by the algorithms selecting news feed stories, targeting advertisements, and recommending new friend connections. As users are more likely to interact with content suggested to them by Facebook, the extent to which these algorithms affect users’ behavior cannot be known fully.

Furthermore, profile entries such as job position or political views are recorded using open-text input fields, which can be equipped with autocomplete features that suggest entries based on the first few letters typed in. This may minimize spelling errors, but it can also introduce biases. For example, a user might have intended to identify himself as a “social scientist,” but may settle on “social scientist” if Facebook suggests the latter. Finally, Facebook’s constant evolution as a platform also drives changes in users’ behavior. Status updates, for instance, were originally preceded with the user’s name (i.e., “Alice Miller is...”), encouraging updates written in the third person. Nowadays, status updates are free to contain any text.

Part #3: Collecting self-reports on Facebook

Collecting self-reports from Facebook users is similar to collecting self-reports in other online environments. In fact, an existing online survey or questionnaire can be integrated easily with Facebook by adding a fragment of HTML code. Obtaining access to participants’ Facebook profiles means that many of the typical questions (such as those concerning demographics) can be skipped given that data can be obtained directly from the Facebook profile or inferred from the targeting approach used to promote the link to the study.

As in other contexts, offering appropriate incentives to participants is an important consideration while designing a Facebook-based study. In general, we discourage financial incentives, as they do not reward people for responding honestly or behaving naturally, but merely for participating in the study. Thus, not only are financial incentives expensive, but they may also encourage dishonest or random responding, and attract semi-professional participants. Rewarding participants with an enjoyable experience or interesting feedback can achieve a much better alignment of the participants’ and researchers’ interests. Even the studies that do not produce interesting feedback could be enriched with elements that do.

Another widespread mistake is to design the study around the needs of the researcher while disregarding the participants’ experience. For example, participants are often prevented from skipping questions or tasks, or barred from accessing the study based on their demographics. Such an approach may work in controlled laboratory settings, but in an online context, it is
likely to trigger dishonest responses or behaviors. In practice, it is usually easier to remove participants and protocols that do not meet the criteria before the analysis, rather than trying to filter them out at the data-collection stage.

Similar problems arise with overly long studies that lead to inattentive responding and high dropout rates. Instead, participants could be engaged gradually by distributing incentives across the study. In the case of the myPersonality project, respondents could choose to receive their reward (feedback on their personality) after answering as few as 20 out of 100 personality questions. This kept the barrier to participation low while encouraging many participants to answer more questions or proceed to other questionnaires.

While our experience indicates that Facebook samples produce self-reported data of very high quality, several issues should be considered. First, the ease of accessing and responding to a study as well as the instant feedback that it offers may encourage participants to rush through it. Second, researchers have little or no control over the circumstances in which a participant is accessing a study, so it is possible that some participants simultaneously engage in other activities. Third, the lack of face-to-face contact increases the psychological distance between the researcher(s) and participants, which may decrease the participants’ feeling of accountability. Finally, because Facebook participants come from diverse backgrounds, they may misunderstand instructions or test questions due to linguistic or cultural differences.

Part #4: Ethical considerations

Unfortunately, there are no clear guidelines on using Facebook or other social media platforms for research. Facebook offers participants a relatively high degree of control over their data, but it is the researcher’s responsibility to weigh the costs and benefits of collecting and using personal user information — and to defer to an IRB when in doubt. The mere availability of data and participants’ willingness to share them does not grant researchers the right to record and use them freely.

The lack of formal guidelines is exacerbated by ever-accelerating technological progress; both researchers and IRB members may over or underestimate the threats to participants, thereby hindering benign projects or approving malignant ones. Both factors discourage social scientists from conducting online research or submitting studies for review. As a result, computer scientists — who are often unconcerned about or unfamiliar with the ethical and social implications of human subject research — are carrying out an increasing proportion of these studies.

This trend is disconcerting, and not only because Facebook constitutes a powerful research tool and an important area of interest for social sciences. We hope to encourage IRBs; federal agencies such as the U.S. Department of Health and Human Services; and the APA Ethics Committee to increase their focus on new research tools and environments, including Facebook. Furthermore, we think that papers employing Facebook data should include a discussion of ethical considerations related to the design of a study and its findings. Such an approach ensures that the authors have considered the ethical aspects of their own work, and supports the evolution of standards and norms in this quickly changing technological environment.

There are two major ethical challenges pertaining to data collection in the Facebook environment. First, the boundary between data belonging solely to participants and information belonging to others is very vague. Participants’ consent allows researchers to record content that refers to or was contributed by other people, such as tagged pictures, videos, messages, or comments on the participant’s profile. In our view, it is acceptable to use data generated by or containing references to nonparticipants, but only if the analyses are aimed exclusively at those directly participating in the study. For example, non-participants’ demographic profiles and network connections could be used to establish the parameters of a participant’s egocentric social networks, or the gender ratio among their friends, but not to study any of the non-participating friends.

The second major challenge refers to the vague boundary between public and private information. Some basic profile information is publicly available and even indexed by search engines. However, some scholars point out that the border between public and private is not determined by accessibility, but by social norms and practices. For instance, in a small town where everyone knows intimate details about everyone else, people tend to pretend not to know facts that are considered personal. Others argue that mining public data is equivalent to conducting archival research, a method frequently employed in disciplines such as history, art criticism and literature, which rarely involve rules for the protection of human subjects.

We lean toward the latter argument and believe that public Facebook profile data may be used without participants’ consent if it is reasonable to assume that the data were knowingly made public by the individuals. Researchers should, however, immediately and irreversibly anonymize the data and abstain from any communication or interaction with the individuals in the sample. Furthermore, researchers should be cautious not to reveal any information that could be attributed to a single individual (such as photographs or samples of text) while publishing the results of the study.

Conclusions

Digital devices and services such as Facebook now mediate a growing proportion of human activities. Social interactions, entertainment, shopping and other information can easily be recorded and analyzed, fueling the emergence of computational social science and facilitating the transition from small-scale
experiments and observational studies to large-scale projects based on thousands or millions of individuals. Observing or experimenting with large samples enables scientists to minimize the problem of sampling errors, which are typical to social science, and to detect patterns that might not be apparent in smaller samples. It also offers unprecedented insights into the dynamics and organization of individual behavior and social systems, with the potential to radically improve our understanding of human psychology.

However, researching psychological phenomena in the digital environment requires skills that are relatively uncommon among social scientists, such as recording, storing, processing and analyzing large databases. Since social scientists have been relatively slow to embrace the skills needed to conduct research using Facebook and similar platforms, data-driven human subjects research is increasingly ceded to computer scientists and engineers, who often lack the appropriate theoretical background and ethical standards.

We strongly encourage our fellow social scientists to not only train themselves in modern computational methods, but to immerse themselves in new human environments, including Facebook. These digital arenas offer new opportunities for social science research and new challenges for researchers in their own right. Additionally, with proper training, psychologists and others can conduct studies at a lower cost and larger scale than ever before.

Michal Kosinski, PhD, is an assistant professor in organizational behavior at Stanford Graduate School of Business.
Sandra C. Matz is a PhD candidate in psychology at the University of Cambridge.
Samuel D. Gosling, PhD, is a personality/social psychologist at the University of Texas at Austin.
Vesselin Popov, is the development strategist for the University of Cambridge Psychometrics Centre, a multidisciplinary research institute specializing in online behavior and psychological assessment.
David Stillwell, PhD, is the deputy director of The Psychometrics Centre at the University of Cambridge.

This article is a condensed version of “Facebook as a Research Tool for the Social Sciences,” published in the September 2015 issue of American Psychologist. To read the full article, which includes all citations and references, go to http://psycnet.apa.org/journals/amp/70/6/543.