**Nikola Banovic** *Carnegie Mellon University, Pittsburgh, PA*

**Editor: Khai Truong**



Photo, veer.com

# TO REPLICATE OR NOT TO REPLICATE?

To replicate or not to replicate experiments and studies seems to be a burning question in the Human-Computer Interaction (HCI) community. This question is equally relevant to the field of mobile computing, where researchers face unique challenges when attempting to replicate studies of mobile device usage in the field. Some of those challenges are inherent in conducting such studies to understand complex socio-technical systems, which involve different technologies and diverse populations of users from different social, economic, and cultural backgrounds. However, those same challenges point to a great opportunity to expand our knowledge of how people use and interact with their mobile devices and understand how those interactions evolve over time.

Replicability is the ability to reproduce an experiment in its entirety and obtain the same results, thus confirming the original findings. To replicate a mobile device usage study would mean to recruit a sample from the same participant population as the original study, use the same method to measure and observe the same phenomenon, perform the same analysis on the collected data, and hopefully find the same results and make the same conclusions as the original study. Replicating mobile usage studies strengthens the original results and provides more evidence that they hold for a particular participant population.

The ability to reproduce an experiment, also called reproducibility, ensures that the original researchers or others can study the same or similar phenomenon. It is important to note a subtle difference between replication and reproduction: replication does not allow for changes in the experiment, whereas reproduction does [8]. For example, researchers can still reproduce a mobile usage study, but conduct the study on another population of users, collect additional data or perform different kind of analysis to obtain new knowledge about the same or similar phenomenon. Thus, every replication is a reproduction, but not every reproduction is a replication.

Researchers who conduct mobile device usage studies in the field often face challenges when trying to replicate and reproduce such studies. Traditionally, when researchers replicate studies, it is with the assumption that all parameters from the previous study are carried over. This is easier to do when conducting studies in the lab than when conducting them in the field, because researchers have more control over the environment [4]. Unlike in some other science disciplines, when conducting mobile usage studies in the field, researchers face highly dynamic and fast evolving environments that change rapidly, making it difficult to study a phenomenon that may have already evolved.

Researchers also face logistical challenges when replicating mobile studies. They depend on the amount of detail the original study authors decided to include in their papers. Also, not all researchers have the same resources at their disposal and access to the same large participant pools to be able to reproduce studies exactly. Researchers also have difficulty publishing replication studies in the mobile computing community [16], primarily because of the community's interest in novelty of contributions in its three core disciplines: Science, Engineering, and Design.

In this article, we argue for reproducing mobile device usage field studies. We argue that the very challenges of replicating mobile studies in this volatile space represent opportunities to frequently take the pulse on this space by reproducing studies at different times and with different populations of users. We then point to recent advances in the mobile space that are helping level the playing field for researchers all over the globe, and enabling them to reproduce studies more easily than before. We discuss why reproducibility is an important hallmark of the science when conducting mobile device usage studies in the field that helps expand our knowledge of how people use and interact with their mobile devices.



Photo, istockphotot.com

## THE CHALLENGE OF REPLICATING MOBILE USAGE STUDIES IN THE FIELD

Replicating mobile usage studies in the field differs greatly from replicating experiments and studies in other sciences, such as, for example, Physics and Chemistry. In those other sciences, experiments are often conducted in the lab, where the environments are highly controlled and all parameters of the original study can be carried over. Repeating the same experiment in those sciences increases the confidence in the results. However, this is very difficult, if not impossible, to do when examining mobile computing interactions and experiences in the field. This is partly because mobile technologies are evolving quickly and people co-evolve with those technologies.

In a very short period of time, mobile devices have evolved from simple communication devices to general-purpose computing devices supporting a variety of mobile tasks. Most mobile device manufacturers have at least one or two release dates per year, and the market expectation is that there will be significant improvements in mobile interactions between those releases. Also, it is common for mobile device manufacturers and application developers to implement cutting edge research and integrate it into mainstream products.

Mobile users also co-evolve with technology and change their behaviors in response to new experiences with those technologies. For example, it would be very difficult to replicate early studies of how people enter text with mini-QWERTY keyboards on touch-screen devices. This is because now there are few people who are novice typists or have little experience

typing on such devices. Even if such people exist, they likely have had more exposure to touch-screen technologies than people in the original studies when such technologies were not as pervasive. Because people may have more exposure to such technologies today than before, their mental models of how these technologies work likely evolved beyond those from the past. This change in understanding how technology works would likely impact the results of replicated studies.

The environments in which people use their devices change, too. Furthermore, researchers have little or no control over such external factors. For example, to replicate Dearman et al.'s [7] paper on how people respond to errors in mobile navigation applications would require finding the same technology used in the study, recruiting participants from the same population, conducting the study at the same time of day and same day of the year, and at the same location. However, there are no guarantees that the study will be fully replicable. The general area of the city where the study was conducted has changed with new buildings introducing new location reading errors.

Even if the field was static, recruiting participants from a particular population of people who participated in the original study would still be difficult. Researchers and their participants are often spread out all over the globe. For example, Ha et al. [12] conducted a study of addiction to mobile devices in Korea. To replicate that study, researchers from other parts of the world may struggle to recruit participants from Korea. Instead, they may be forced to reproduce the study in their own country. Even then, it is likely that the findings of the two studies would differ in part because of the cultural and social differences between people in the two countries.

It is therefore not reasonable to expect that replicating mobile studies in the field will yield the same results from study to study. Every mobile device usage study in the field is taking a snapshot of the current behaviors of people, in current environments, and with currently available mobile interactions and applications. When only some parameters of the study can or need to be fully replicated, researchers still have an option to reproduce the study instead of trying to fully replicate it. In fact, reproducing mobile studies in the field allows the researchers to keep track of how technology, people, and environments evolve over time.

**EVERY REPRODUCTION OF A MOBILE USAGE STUDY AT A DIFFERENT TIME, AND WITH USERS FROM DIFFERENT SOCIAL, CULTURAL, AND ECONOMIC BACKGROUNDS EXPANDS OUR KNOWLEDGE OF THIS DYNAMIC FIELD.**

**SUCH KNOWLEDGE DRIVES THE INNOVATION IN THE FIELD OF MOBILE COMPUTING.**

## TOWARDS REPRODUCING MOBILE USAGE STUDIES IN THE FIELD

Reproducing mobile studies is a way to enable researchers to re-examine the same phenomenon at different points in time. Note that the goal of reproduction studies is not necessarily to try to look for differences or to establish causality between different factors in a study. Instead, the reproducibility of mobile usage studies focuses on the ability of researchers to repeat a study of a phenomenon.

For example, to learn whether same mobile device usage trends hold over time in response to always changing mobile trends, it is important to keep track of those frequent changes through reproducing mobile device usage studies in the field. This is important because most studies that propose new mobile interactions could be just a release cycle away from being mainstream. For example, in 2013 a series of studies [1, 13, 14] explored different ways in which users can interact with mobile device lock screens – interactions not available on any platforms at the time. Around the time of the publication of those studies in 2014, Google launched Android Lollipop that made interactive lock screen widgets part of its operating system. Reproducing those studies now would shed light on how users adopted the interactions, and how their behavior differs from participants in the original studies in response to those changes.

Another reason it is beneficial to reproduce mobile usage studies is to confirm whether findings of original studies generalize to other populations of users. For example, in a recent paper, Church et al. [6] reproduced the same study analysis of short mobile device usage from Banovic et al. [1] and short application usage from Ferreira et al. [10] on three different data sets. The three data sets contained usage data from security minded Android users [11], Android users who used Google or Yahoo search on their mobile devices [5], and teen users from the San Francisco Bay Area. The stark differences between mobile device usage of participants in each of the three data sets and the original two studies show that little of the original findings can generalize across those different user groups. Such findings also raise questions whether differences exist within seemingly similar populations (e.g., findings from the teen study do not necessarily generalize to teens in other parts of the world). Therefore, reproducing studies also helps discover the differences within and between different populations of users.

However, every change from the original study has implications on the ability to compare the findings from the new study against the original findings. Therefore, any changes that researchers make must be based on the goals of the reproduction. For example, to study whether findings from past studies about how people use mobile technology still hold today would require not changing the values of the

parameters from the original study that researchers have control over. Similarly, to verify that findings from an original study generalize to a different population requires keeping all parameters of the original study the same except for the new population. This equates the reproduction effort to replication with the understanding that factors that researchers have no control over might have changed.

Because reproduction studies should still make an effort to keep important parameters of the original study unchanged, some of the challenges of replicating studies also exist when reproducing studies. For example, to properly reproduce a study requires detailed description of the original study, which might not always be available. Also, not all researchers have the same resources to conduct their studies in the same way or compensate their participants with the same incentives. However, this represents an opportunity for the mobile computing community to find ways to make it easier for researchers all over the globe to reproduce mobile usage studies and expand our knowledge of different user populations around the world.




## LEVELING THE PLAYING FIELD

In recent years, it has become easier to reproduce existing mobile usage studies in the field. Publicly available mobile usage tracking applications and frameworks make it possible to use the same study software across different studies. For example, Aware Framework comes with a set of modules that track core sensors on the Android platform, but also enables researchers to develop and share their own plugins that extend the tracking capabilities of the framework. Researchers are also likely to share the source code for their study software and mobile interventions on platforms such as GitHub.

Application stores, such as Google Play Store and Apple's App Store, enable researchers to distribute their study applications to large and diverse user population samples. For example, Böhmer et al. [3] recruited 4,125 participants through an application store and studied their mobile application usage. Wagner et al. [15] posted their tracking application on Google Play Store and collected mobile usage data in the field from over 12,500 participants over a period of two years. Posting study applications on the application stores enables researchers to reach participants from all over the globe and to target participants with the same social, economic and cultural backgrounds as the original studies.

To recruit a large number of participants without spending large amounts of money, researchers have started to offer different incentives to their participants instead of monetary compensation. One such incentive is to provide participants with applications and services that they need. Examples include a free battery monitoring application, which was downloaded and used by 4,437 participants [9], and a mobile application that tells participants about security threats on their phones which was installed and used by over 200 participants [11]. Also, Aware Framework now offers participants the ability to track their own mobile usage from many sensors in exchange for sharing the data with researchers. Such incentives allow researchers with fewer resources to reproduce the same studies in the field without having to spend money to compensate their participants.

The above advances continue to improve the reproducibility of mobile usage studies in the field, but some challenges remain. For example, just as the success of an original study application posted on an application store depends on many factors [2], so does the success of reproducing the study with the same application. If the original application is still available on the application store, it might be difficult to post another similar or the same application and expect the same download rates. Another major barrier to reproducing studies is that the existing mobile usage tracking frameworks do not work across different platforms (e.g., Android, iOS). However, being able to test the generalizability of findings from one mobile platform to another depends on the ability to reproduce the same study on a different platform.

To properly reproduce a study without adding any additional uncertainty and unnecessary variations it is important to have a detailed description of the method and the procedures of the study. Church et al. [6] reviewed mobile usage studies conducted in the past ten years and found that the papers reported high-level information about the participants, method, and the study software. However, this information may not always include all the detail required to fully reproduce the study. Ideally, the authors should share their study protocols and their complete study and analysis software.

To address the challenges described above and improve reproducibility of studies, the mobile computing community will have to push for reproducibility as a hallmark of science in mobile computing. To understand how to best replicate or reproduce a study, mobile computing community will have to document the success and failure stories of different attempts. Analysis of those case studies could lead to methodologies that can help improve the success of replicating mobile usage studies in the field.

## VIEWING REPRODUCIBILITY AS A HALLMARK OF SCIENCE

Another great challenge that researchers face when reproducing studies is difficulty publishing their results [16]. One of the main barriers to publishing such studies is a general interest in novelty in the three

core disciplines within mobile computing: Science, Engineering and Design. Science in mobile computing seems to be rushing towards new discoveries and knowledge that helps in understanding how people interact with technology. Engineering in mobile computing seeks to make technology that goes faster, stores more information, or enables new communication that was not possible before. Design strives for novel interactions, experiences, and services that improve, change or disrupt status quo.

However, as we have argued above, every reproduction of a mobile usage study at a different time, and with users from different social, cultural, and economic backgrounds, expands our knowledge of this dynamic field. Such knowledge drives innovation in the field of mobile computing. The scientific contribution of reproducing mobile usage studies comes from the new knowledge these studies provide about people's current usage behaviors. Furthermore, reproducing the same studies at different times can help scientists find trends in mobile device usage behaviors and build models that help predict future trends. Reproduction studies help identify recurring interaction bottlenecks and inform which technologies engineers in mobile computing should focus on improving. Designers in mobile computing work to understand and change complex social-technical systems that evolve quickly. To design novel mobile interactions that improve on the existing experience with mobile devices, it is necessary to explore and understand the most recent "pain points" in the field. Reproducing studies also allows designers to evaluate the adoptability of their designs and the impact of their interventions over time.

This article is therefore a call to the whole mobile computing community to establish reproducibility as an important part of studying mobile device usage in the field. As researchers reproduce studies, we need to understand and argue for the benefits of this important aspect of the scientific method. We must encourage academic reviewers to consider how results of reproduction studies contribute to the larger body of knowledge of the field. Most importantly, this is a call to the mobile research community to help improve the study reproduction methods and make them accessible to a wider audience of researchers. Establishing reproducability as a hallmark of science in mobile computing will improve the methodology of mobile studies in the field. ■

**Nikola Banovic** is a PhD student at the Human-Computer Interaction Institute (HCII) at Carnegie Mellon University. His research focuses on modeling and understanding human routine behaviors, including routines of mobile device users.

## REFERENCES

[1] Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. Proactive Tasks: the short of mobile device use sessions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 243-252.

[2] Susanne Boll, Niels Henze, Martin Pielot, Benjamin Poppinga, and Torben Schinke. 2011. My App is an Experiment: Experience from User Studies in Mobile App Stores. *Int. J. Mob. Hum. Comput. Interact.* 3, 4 (October 2011), 71-91.

[3] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling asleep with Angry Birds, Facebook and Kindle: a large-scale study on mobile application usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (MobileHCI '11). ACM, New York, NY, USA, 47-56.

[4] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportuni-ties for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 1657-1666.

[5] Juan Pablo Carrascal and Karen Church. 2015. An In-Situ Study of Mobile App & Mobile Search Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 2739-2748.

[6] Karen Church, Denzil Ferreira, Nikola Banovic, and Kent Lyons. 2015. Understanding the Challenges of Mobile Phone Usage Data. In P*roceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '15). ACM, New York, NY, USA, 504-514.

[7] David Dearman, Alex Varshavsky, Eyal De Lara, and Khai N. Truong. 2007. An exploration of location error estimation. In *Proceedings of the 9th international conference on Ubiquitous computing* (UbiComp '07), John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang (Eds.). Springer-Verlag, Berlin, Heidelberg, 181-198.

[8] Chris Drummond. 2009. Replicability is not reproducibility: nor is it good science. In P*roceedings of the Evaluation Methods for Machine Learning Workshop at ICML.*

[9] Denzil Ferreira, Eija Ferreira, Jorge Goncalves, Vassilis Kostakos, and Anind K. Dey. 2013. Revisiting human-battery interaction with an interactive battery interface. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (UbiComp '13. ACM, New York, NY, USA, 563-572.

[10] Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. 2014. Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 91-100.

[11] Denzil Ferreira, Vassilis Kostakos, Alastair R. Beresford, Janne Lindqvist, and Anind K. Dey. 2015. Securacy: an empirical investigation of Android applications' network usage, privacy and security. In P*roceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks* (WiSec '15). ACM, New York, NY, USA, Article 11, 11 pages.

[12] Jee Hyun Ha, Bumsu Chin, Doo-Heum Park, Seung-Ho Ryu, and Jaehak Yu. 2008. Characteristics of excessive cellular phone use in Korean adolescents. In *CyberPsychology & Behavior* 11, no. 6, 783-784.

[13] Khai N. Truong, Thariq Shihipar, and Daniel J. Wigdor. 2014. Slide to X: unlocking the potential of smartphone unlocking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 3635-3644.

[14] Rajan Vaish, Keith Wyngarden, Jingshu Chen, Brandon Cheung, and Michael S. Bernstein. 2014. Twitch crowdsourcing: crowd contributions in short bursts of time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 3645-3654.

[15] Daniel T. Wagner, Andrew Rice, and Alastair R. Beresford. 2014. Device analyzer: large-scale mobile data collection. *SIGMETRICS* Perform. Eval. Rev. 41, 4 (April 2014), 53-56.

[16] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI should be replicating and validating results more: discuss. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '11). ACM, New York, NY, USA, 463-466.