

How Carat Affects User Behavior: Implications for Mobile Battery Awareness Applications

Kumaripaba Athukorala ‡§

Antti Jylhä ‡§

Eemil Lagerspetz ‡§

Adam J. Oliner †

Giulio Jacucci ¶§

Maria von Kügelgen ‡§

Sasu Tarkoma ‡§

‡ Helsinki Institute for Information Technology HIIT, University of Helsinki, † UC Berkeley

§ Department of Computer Science, University of Helsinki

¶ Helsinki Institute for Information Technology HIIT, Aalto University

§ first.last@helsinki.fi, † oliner@eecs.berkeley.edu

ABSTRACT

Mobile devices have limited battery life, and numerous battery management applications are available that aim to improve it. This paper examines a large-scale mobile battery awareness application, called Carat, to see how it changes user behavior with long-term use. We conducted a survey of current Carat Android users and analyzed their interaction logs. The results show that long-term Carat users save more battery, charge their devices less often, learn to manage their battery with less help from Carat, have a better understanding of how Carat works, and may enjoy competing against other users. Based on these findings, we propose a set of guidelines for mobile battery awareness applications: battery awareness applications should make the reasoning behind their recommendations understandable to the user, be tailored to retain long-term users, take the audience into account when formulating feedback, and distinguish third-party and system applications.

Author Keywords

user retention; user behavior; smartphone; energy awareness

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; H.1.2. User/Machine Systems: Human Factors

INTRODUCTION

Mobile devices have limited battery life, sometimes requiring a recharge more than once per day. Rapid energy drain may be caused by extensive use of resources (e.g., the network, CPU, or GPS) by running applications or the device operating system, itself. Poor battery life contributes negatively to user experience [21].

To automatically increase battery life or help users manage power consumption (a.k.a. *battery awareness*), numerous battery-saving applications have come to application markets [4, 15, 16, 25]. Although these applications can improve battery life, the automated solutions do not typically give users a direct indication of the concrete actions that make the battery last longer. Such applications, therefore, tend not to guide user behavior towards battery-saving choices.

There is prior work on the effect on user behavior of household energy awareness applications [1, 3] and mobile battery level indicators [20, 22]. However, we are not aware of any user behavior studies in the context of mobile battery awareness applications.

In this paper, we examine users of Carat, a community-based mobile battery-awareness application deployed worldwide to more than 670,000 devices. We conducted a survey of over 1,000 Carat users and analyze their responses along with data automatically gathered by Carat. Prior work on the Carat logs has shown not only that the application recommendations improve battery life (11% after 10 days and 40% after 90 days, on average), but that there is a positive correlation between the duration of using the application and the extent of the improvement [15]. One question we examine in this work is what distinguishes these long-term users from short-term users that might explain the difference in battery life improvement.

The contributions of this work are as follows:

- Elucidates the relationship between mobile battery awareness applications and user behavior;
- Examines two classes of users, distinguished by duration of use of the application, with distinct behaviors;
- Articulates lessons learned in the form of actionable guidelines for future battery awareness applications.

RELATED WORK

There is a growing body of literature on improving the battery life of mobile devices. The majority of this work consists of technical solutions, not intended for novice users [12, 16]. Other work provides suggestions or guidelines to users on how to reduce battery consumption [6, 14, 15]. There is another line of research on Human-Battery Interaction (HBI)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI'14, April 26–May 1, 2014, Toronto, ON, Canada.

Copyright © 2014 ACM 978-1-4503-2473-1/14/04\$15.00.

<http://dx.doi.org/10.1145/2556288.2557271>

that focuses on user interaction with different battery indicators. Understanding the battery charging behavior and users' knowledge of power-saving features has been the subject of some HBI studies [2, 7, 20, 22]. In the domain of household energy awareness, there has been research on user behavior changes [1, 3]. We are not aware of any user behavior studies in the context of mobile battery awareness applications. As the goal of this research is to understand how mobile battery awareness applications change behavior, this section mainly considers work related to such applications, human-battery interaction, and behavior studies in the domain of household energy awareness.

Mobile Battery Awareness Applications

The primary goal of mobile battery awareness applications is to make the user aware of what consumes energy. The Android operating system has a built-in energy profiler that shows statistics about battery use on the device. This can be accessed from the battery option in the settings on most devices. Early consumer tools for energy awareness on smartphones include the Nokia Energy Profiler [4], which runs in the background, recording phone subsystem use, and later reports the energy use (in watts) over time. A more recent profiler is PowerTutor for Android [25], which shows energy use similarly to Android's built-in profiler but broken down by resource (e.g., CPU, WiFi, and the screen) and by category (e.g., by application or system component).

Carat is the first collaborative approach to mobile battery awareness, which allows it to perform diagnoses which would be impossible on a single device [15]. For example, although the tools discussed in this section are able to obtain accurate energy consumption profiles on a device, they cannot determine whether the amount of energy used by an application or device is normal. With a community of hundreds of thousands of client devices, Carat identified applications that consume abnormally large amounts of energy compared with other applications as well as instances of individual applications that consume abnormally large amounts of energy on only a subset of devices.

Human Battery Interaction (HBI)

Some work in HBI considers how users deal with limited mobile battery life. Banerjee et al. [2] studied phone battery-charging behavior and identified two categories of users: those who charge their phones regularly regardless of the remaining battery level and those who charge based on battery status indicators. Furthermore, there was usually a significant amount of power left in the battery when it was charged, even for users who charge based on the battery status indicator. There is also research on how battery-use feedback in mobile phones affects behavior [20, 22]. Studies revealed that some battery charging habits can reduce battery lifetime [7]. We extend these lines of inquiry by studying how mobile battery awareness applications affect behavior.

Behavior Change in Energy Awareness Applications

Literature on feedback for energy conservation spans several decades and includes work from several disciplines, including the behavioral sciences. This literature mostly considers

domestic settings [18]. Although battery awareness applications address a different problem from domestic energy consumption feedback, there are similarities. One key difference for mobile battery management is that power conservation is motivated by extending use time, while in domestic settings motivations are environmental or monetary.

Carat follows some of the key principles proposed in the energy consumption feedback literature; it provides actionable feedback, rewards users to keep them motivated, and avoids information overload. A solution tailored to individual users facilitates the job of persuading users to take the suggested actions, as each behavior has its own personalized reasons and constraints [1, 3].

Effective feedback should be real-time [3] and actionable, demonstrating a way to fill the gap between current actions and desired goal state [13]. The goal should always be clear to the user and be accompanied by instructions on how to achieve it.

Sustained involvement requires interfaces that evolve, rewarding improvements to keep the user motivated after the initial curiosity drops [11]. Recent work proposed a three-stage approach for feedback including the following: raise awareness, inform complex changes, and maintain sustainable routines [23].

There are a variety of other principles (e.g., format of feedback [9, 17, 19]), but the ones above are the most relevant for the case of battery awareness applications. Most available applications only provide simple feedback [23] ranging from power measures to monetary charges to carbon footprints. By themselves, such numbers do not suggest clear actions to take. Like recent energy consumption solutions that include contextually triggered advice [10], Carat provides concrete, actionable suggestions to users to improve battery life.

INTRODUCTION TO CARAT

Carat [15] uses a collaborative black-box method for diagnosing anomalies on mobile devices. Carat is an application on both the iOS and Android platforms.

The client application sends intermittent, coarse-grained measurements to a server. The server correlates running applications, device model, operating system, and other features with energy use. The system generates actions that the user could take to improve battery life. The amount of improvement error, and confidence of the suggestions given by Carat is presented to the user along with the actions. Carat has been installed on more than 670,000 devices.

On a single device, it is not possible to diagnose all types of abnormal energy use, because it could result from device or user-specific factors. A collaborative approach is required to diagnose energy bugs of this kind. Carat achieves this by using a community of devices.

Carat in Action

To walk through the features of Carat, we use the following scenario. John is a smartphone user with battery life issues. He starts by installing Carat on his phone. At first, when John

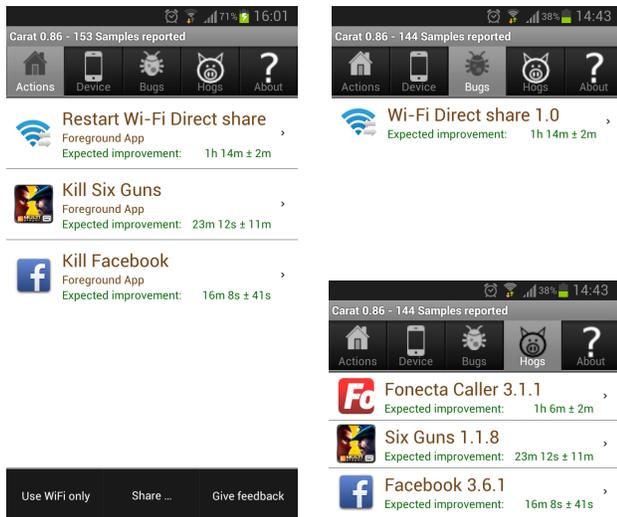


Figure 1. The Actions, Hogs, and Bugs screens

opens Carat, the system has no data on his device. However, he may see suggestions based on his currently running applications that are known anomalous energy consumers in the Carat community.

To calculate results for John’s phone, Carat needs data from that phone. Carat application gathers intermittent measurements of the running applications, battery level, and other device features. When John opens Carat, it sends these data, referred to as samples, gathered so far to the server. John needs to open Carat regularly, preferably at least once a day, to receive personalized results as quickly as possible. He can see the number of sent samples in the top bar of the Carat application (Figure 1).

Actions

After about a week, John receives his first results. On the opening screen of Carat, the “Actions” tab, he sees suggestions given by Carat (Figure 1). There are usually two types of suggestions: “Kill application X” or “Restart application Y”. Carat also shows how much the battery life is expected to increase if these applications were not used. If John wishes to kill or restart these applications, he can click on the corresponding item, and Carat will show a screen with instructions on how the application can be killed. Most of the time, John can kill an application just by clicking a button on this Carat screen, and Carat provides alternative instructions in case that fails. Sometimes, killing an application on Android does not succeed, because the application’s background service restarts it right after it has been killed. Applications that behave like this can be *force killed* through the Task Manager, accessible from the same screen of Carat. However, doing this for applications required by the system can lead to instability, and often there is not much the user can do about them.

Bugs and Hogs

John can also see the applications mentioned in the “Actions” screen on the “Bugs” and “Hogs” tabs (See Figure 1). There might be other applications listed in those tabs as well, since

actions are suggested only for the applications that are running on the users phone at that moment. On these tabs John sees lists of applications that he has been using after installing Carat that have been classified as *bugs* or *hogs*.

Hogs are applications that use more energy than an average application in the Carat community. Typically hogs require more energy for normal function; examples of this are VoIP, Internet radio, navigation, and camera applications. However, hogs can also result from a widespread problem with an application’s energy use.

If the application is not a hog, it can still be a bug. A bug is an application that, for some reason, uses more battery than average on a specific device. For example, the Kindle application uses less energy, on average, than the average application, so it is not a hog. However, as reported in [15], some versions of the Kindle application had a bug which caused it to use more energy when connected via a mobile network. This made Kindle show up as a bug for Carat users who preferred mobile networks over WiFi. How hogs and bugs are calculated is described in detail by Oliner et al. [15].

The actions along with the hog and bug reports help John understand which applications are draining the battery faster than others, but also which of these applications are often running when the user opens the Carat application. He can use the actions screen to kill or restart running applications, and from hogs and bugs screens he can gain wider knowledge about applications that lower his battery lifetime and that of the other users in the Carat community.

J-Score and Other Information about Battery Life

After getting to know the energy efficiency of his applications better, John starts to get more interested in how well his battery lasts compared to other people. On the “Device” tab (see Figure 2) he sees a value called the J-Score. The J-Score tells him the percentage of devices in the Carat community that have a worse battery life than his phone. Underneath the J-Score is the expected active battery life calculated by Carat. This shows how long the battery would last if the device was used in the way that John has been using it since he started using Carat.

Other Features of Carat

John also sees some basic information about his phone on the “Device” tab: the operating system version, the device model, and information about the memory use of his phone. He can also look at a list of all the currently running applications by clicking the button “View Process List”.

METHOD

We collected data from the existing Carat users by following two data collection methods: survey and system logs of Carat.

User Survey

Our first goal was to identify how different features of Carat affect user behavior. Therefore, we constructed a questionnaire and placed a link to it on the opening screen of Carat. The link was published to all Carat Android users. The survey was open for two weeks starting from August 12th, 2013.

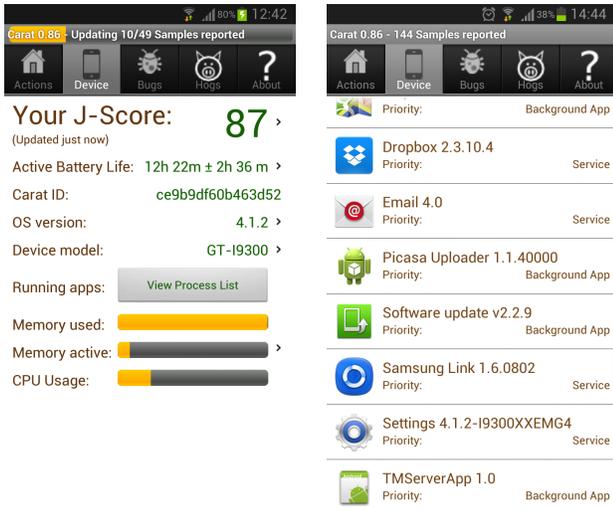


Figure 2. The Device screen and the process list

Survey Questionnaire

The questionnaire was composed of 16 questions plus optional free-text fields to express any additional comments regarding Carat application use experience and suggestions for improvement. All the questions were multiple choice, except for questions 8, 9, and 12, which were 7-point Likert scale. The full questionnaire with the multiple choice answer options can be found on the Carat website¹. The following questions were asked in the questionnaire:

1. How long have you been using Carat?
2. What kind of device are you using now?
3. Do you use external batteries for this device?
4. How often do you charge the battery on this device?
5. On how many devices do you use Carat?
6. What is your main reason for using Carat?
7. Why did you choose Carat (instead of some other energy saving app)?
8. How well do you understand how Carat works?
9. Are you interested in knowing how Carat works?
10. What is the main reason for opening the Carat app?
11. Which of the following things do you do most times when you open Carat?
12. How often do you kill or restart an app when Carat suggests it?
13. What are the reasons why you don't kill an app when Carat suggests it?
14. How often have you opened Carat during the past month?
15. In what kind of situation do you usually open Carat?
16. In what ways has using Carat changed the way you use your device?

Response Statistics

A total of 1,140 valid responses were received from dozens of countries covering many of the regions with Carat users. Among the respondents 16% had been using Carat for over a year, while 40% had been using Carat for less than three

¹<http://carat.cs.berkeley.edu/chi/Carat-usage.html>

Samples	Reports
user id	average battery life (h)
date and time	hogs, the date they were found
battery level	bugs, the date they were found
running applications	list of applications the user has run

Table 1. Contents of Carat logs.

months. Most of the respondents (93%) had been using Carat on their mobile devices while few of them had been using Carat on tablet devices. Around 26% of the respondents had been using Carat on more than one mobile device. 89% of the respondents were male, and the average age of respondents was 37 years. We are aware of the limitations of self-reporting, and we discuss them in the Limitations section.

Carat Logs

In addition to the survey responses we also used automatically gathered Carat usage logs like Carat *samples* and *reports* of the users who answered the survey. The Carat application sends data to the servers in the form of samples. Each sample contains information about application use and battery life. After enough samples have been collected, Carat generates reports about users and applications. These reports are not available on the mobile client. These include details about the user's average battery life, the most battery-consuming applications that have been running on their device (hogs), and any applications that use more energy on their device than in the rest of the community (bugs). In this research we used these samples and reports to quantify user behavior. The contents of the two types of data used in this paper are detailed in Table 1. The logs give us important information, such as when a problematic application was reported to the user by Carat, and how that changed the behavior of the user in terms of running that application.

RESULTS

This section discusses the responses of the questionnaire and results of Carat log data analysis. To quantify differences and correlations in our results, we use two statistical tests. When comparing beginners and advanced users, we apply the Mann-Whitney U test; when discussing correlation, we use Kendall's tau (τ).

Beginners and Advanced Users

Prior work shows that there is a positive correlation between duration of Carat use and battery life [15]. In this paper we examine what are the reasons for this, and what features of Carat and user behaviors cause this positive correlation.

We found a significant positive correlation between the responses to "How long have you been using Carat?" and "How well do you understand how Carat works?" (seven-point Likert Scale where 1 = not at all, 7 = very well), $r_{\tau}(N = 1,140) = .13, p < .001$. There was also a significant positive correlation between the responses "How long have you been using Carat?" and "On how many devices do you use Carat?", $r_{\tau}(N = 1,140) = .26, p < .001$. These results suggest that long-time Carat users believe that they

Definition of beginner	$U(Z)$	p	r
< 1 month	81,027(-4.324)	$p < .01$	-.128
< 3 months	132,180(-4.396)	$p < .01$	-.130
< 6 months	136,352(-3.688)	$p < .01$	-.109
< 1 year	71,478(-4.150)	$p < .01$	-.123

Table 2. Results of Mann-Whitney U test conducted among the two groups of Carat users (classified as beginners according to the given definition) on how well they understand how Carat works. Each classification scheme defines advanced users as those who have used longer than beginners.

Group Characteristics	Advanced Users	Beginners
Number of Respondents	689	451
Duration of Carat Use	>90 days	<90 days
Gender	male = 620 female = 58 other = 11	male = 389 female = 60 other = 2
Mean age in years	38	36

Table 3. Primary characteristics of the two groups of Carat users.

better understand how Carat works and have it installed on many devices.

We compared the responses to the question “How long have you been using Carat?” with the length of Carat usage logs from the device that was used to answer the questionnaire. The two are significantly correlated ($r_{\tau}(N = 1,072) = .357, p < .001$). The Carat logs underestimate actual duration of use because re-installation of the application or migration to a new device is recorded as a new user. The true correlation is therefore likely to be higher. In light of this limitation in the Carat log data, we use the questionnaire responses as a proxy for how long the respondents have used Carat.

Behavioral studies conducted with users of energy awareness applications have found that habits formed over three months are likely to stick with users [5]. We analyzed the survey responses to investigate the validity of this finding in the context of mobile battery awareness applications. First, we classified the respondents as beginners and advanced users by using each of the five options we gave them in the questionnaire (Less than a month, 1-3 months, 3-6 months, 6-12 months, and over year) as the threshold value. Next, for each classification we separately conducted a Mann-Whitney U test on how well they believe they understand how Carat works. A summary of the results is given in Table 2. According to the results, all the four classification schemes result in a significant difference between beginners and advanced users ($p < .01$). However, when we classify those who have been using Carat for less than three months as beginners, the relationship between the duration of use of Carat and how well they believe they understand it is stronger than in others ($r = -.130$). Based on these results and the importance of the three-month milestone in previous work [5], we use that as the classification threshold in this paper. Table 3 summarises the characteristics of these two groups.

Why Users Open Carat

In the survey, we asked the respondents to select from a list of options the main reason for opening Carat. We provided

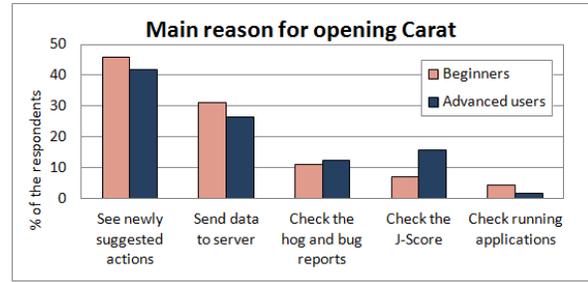


Figure 3. The main reason for opening Carat.

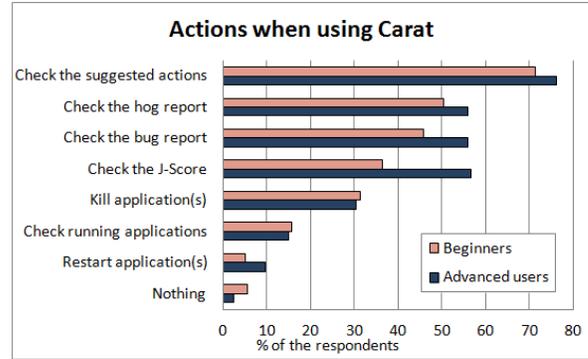


Figure 4. The actions performed when Carat is open.

the primary features of Carat as the options: send data to server, see newly suggested actions (kill or restart an application), check the reports (bugs and hogs), check the J-Score, and check running applications. Figure 3 summarizes the responses.

The majority of respondents (44%) mainly open Carat to see if any actions are suggested for them, and according to Figure 3 it is clear that nearly a similar proportion of beginners and advanced users have selected this reason. Group-wise analysis shows that for both beginners and advanced users, sending data to the server is the second-most-popular reason (27% of all the respondents selected this). Advanced users (16%) were more interested in checking the J-Score than beginners (7%). The respondents were also asked which actions they perform most times when they open Carat (Figure 4). 71% of the beginners and 76% of the advanced users mentioned that they check the suggested actions. About half of all the users check the hog and bug reports, advanced users slightly more often than the beginners. The majority (57%) of the advanced users check the J-Score, but only 36% of the beginners are interested in it. A bit less than a third of the users kill applications most times they open Carat. The running applications are checked by 15% of the users, and applications are restarted by less than 10% of the users. Some users stated that they do nothing most of the time when they open Carat. The percentage of these users is higher among beginners (5.5% compared to 2.5% in the advanced users’ group), probably because Carat does not give results to the user during the first week after installing Carat, so there is not much to do at that point.

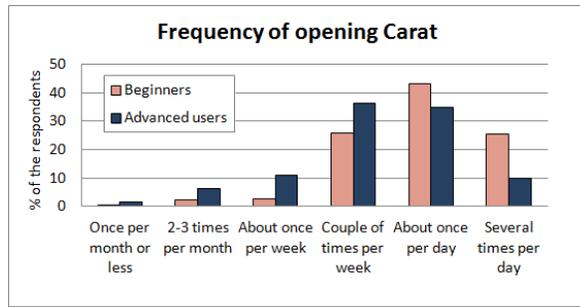


Figure 5. Summary of responses to the survey question “How often have you opened Carat during the past month”.

The primary reason for opening Carat and the most common actions performed after opening Carat do not vary much across beginners and advanced users. However, features like J-Score are more popular among advanced users. Based on these findings, we suggest that advanced users enjoy functions that support comparing against others in the Carat community.

How Often Users Open Carat

We asked the respondents to rate how often they have opened Carat during the past month. Beginners open Carat significantly more frequently than the advanced users. Figure 5 summarizes the responses to this question.

Beginners and advanced Carat users differ in their responses to the question “How often have you opened Carat during the past month?” (Figure 5 contains the options), ($U = 107,132, p < .001, r = .28$). Advanced Carat users had an average rank of 500 (Mdn = 4, SD = 1.08), while beginners had an average rank of 677 (Mdn = 5, SD = .95).

We infer that the suggestions provided by Carat are more useful to the beginners, and over time users learn to manage their battery without repeatedly checking Carat. These findings further suggest that the Carat use behavior changes over time, and the user’s knowledge about how to improve battery life also grows with use of Carat.

Who Follows Suggestions and What They Gain

We asked the respondents to rate how often they kill or restart an application when Carat suggests it (seven-point Likert Scale where 1 = never and 7 = always). Most respondents follow application kill or restart suggestions (mean = 4.39) and beginners and advanced users follow Carat suggestions equally often. However, the users who claim to understand better how Carat works, charge their devices less often and follow Carat suggestions more often.

The difference between beginners (Mdn = 4, SD = 1.91) and advanced Carat users (Mdn = 5, SD = 1.76) on how often they follow Carat suggestions was not significant ($U = 155,203, p = .975, r = .0009$), suggesting that the two groups are similar in how often they kill or restart an application that Carat suggests.

There was a significant negative correlation between the responses to the questions “How well do you understand how

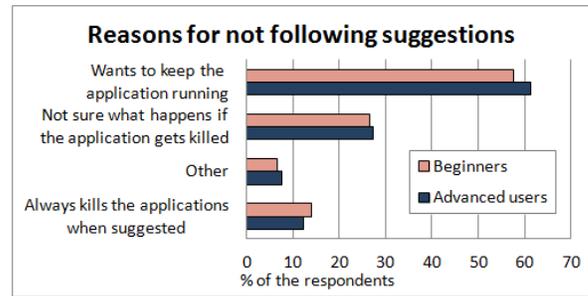


Figure 6. The reasons why users ignore suggestions to kill applications.

Carat works?” and “How often do you charge the battery” ($r_r(N = 1,140) = -.084, p < .01$). Further, we also found a significant positive correlation between how often users kill or restart an application that Carat suggests, and how well user believe that they understand how Carat works ($r_r(N = 1,140) = .071, p < .01$). This suggests that even though the duration of use of Carat does not affect how often users follow Carat suggestions, how well the user understands how Carat works has an effect.

Statistical comparison between beginners and advanced Carat users on the percentage of battery life improvement (collected from Carat logs) after using Carat was significant ($U = 15,200, p < .05, r = -.102$). Advanced Carat users had an average rank of 227 (Mdn = .011, SD = 10.2), while beginners had an average rank of 196 (Mdn = -.53, SD = 10.1). In agreement with prior work [15], we found that duration of use correlates positively with battery life improvement.

All these results suggest that Carat fosters learning, and as a result of that users learn to manage their battery better with long-term use. This encourages the users to stick with Carat. We conclude that energy awareness applications should make the logic behind their suggestions understandable to the users in order to support learning, and encourage them to follow the suggestions and use the application for long.

Why Users Ignore Suggestions

The users of energy awareness applications do not always follow the suggestions provided to them. In order to find why Carat users sometimes ignore the suggestions, we asked the respondents to select from a list of options all the reasons for not killing an application suggested by Carat. The options were: I want to keep it running, I’m not sure what happens if I kill it, I always kill the application when suggested, and a free text field to provide other options. Figure 6 provides a summary of responses. The main reason for ignoring the suggestions to kill applications is that the user wants to keep that application running (58% of the beginners and 61% of the advanced users) regardless of its high power consumption. Some of the respondents provided further justifications for this option in the free text field. According to them, one of the most common reasons that eight beginners and 18 advanced users stated was that some applications cannot be killed. Thirteen advanced users and two beginners stated that they check the estimated battery improvement provided by Carat, and if it’s too low they do not kill the suggested application. Seven

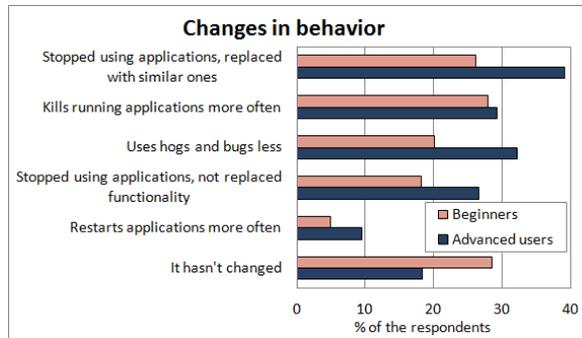


Figure 7. Summary of responses to the survey question “In what ways using Carat has changed the way you use your device?”.

advanced users and one beginner mentioned that sometimes Carat suggests to kill system applications. Eight beginners also stated that they have heard that killing applications in Android is bad.

We conclude that suggestions to kill system applications and regularly used applications are not very useful to the user. However, the estimated battery life improvement number provides additional information for the user to decide whether to kill an application or not.

How Carat Changes Device Use

To understand how Carat has influenced the mobile device use behavior, we asked the respondents to select all the relevant options from a list of user behavior changes that we expected Carat to cause. Responses revealed that Carat has caused behavioral changes especially in advanced Carat users. Figure 7 provides a graphical illustration of the list of given options and the percentage of users selected each option.

Carat did not affect the mobile device use behavior of 29% of the beginners. However, 39% of advanced users agreed that Carat has made them stop using some applications and replace them with similar ones. The second most common change that Carat has caused in 28% of beginners and 29% of advanced users is killing running applications more often. Advanced Carat users agreed on all the behavioral changes more than the beginners.

These results suggest that Carat has a bigger impact on the mobile device use behavior of advanced users, and it takes time for new users to adapt these new habits. This further explains why initial performance analysis indicated that the battery life of Carat users improves gradually over time [15].

Who Kills Bugs, Hogs, and Other Applications

From the Carat logs, we learned that beginners reduced the use of 64.3% of their hogs and bugs on average when they were first reported. All hogs were reduced by 36.5% and bugs by 23% on average. Advanced users reduced the use of 67.2% of their hogs and bugs, hogs by 46% and bugs by 30%. Table 4 shows these reduction ratios.

Furthermore, we compared how much beginners and advanced Carat users reduce the use of hogs and bugs. The difference for hogs was significant ($U = 58,369, p < .001, r = .17$). Advanced Carat users had an average rank of 427 (Mdn = 45, SD = 23) while beginners had an average rank of 345 (Mdn = 36, SD = 25). These results indicate that advanced users reduce the use of hogs significantly more than the beginners. This can be one of the reasons why advanced users improve battery life better than the beginners. However, the test for bugs was not significant ($U = 12,995, p = .189, r = .07$), indicating that both the beginners (Mdn = 27, SD = 38) and advanced users (Mdn = 33, SD = 39) have equally reduced the use of bugs. We also calculated the reduction in use of other applications that have not been reported as hogs or bugs. Here we considered actions, such as starting to use new applications, abandoning old ones, and killing applications for battery saving. Beginners have reduced use of these other applications by 11.51% and advanced users by 24.14%. Since the percentage of reduction in use of other applications was normally distributed, we conducted an independent T-test on this data and found that the advanced users ($M=17.9\%, SD=1.99$) have reduced the use of other applications significantly more than the beginners ($M=5.81\%, SD=1.86$); $t(851)=-8.88, p < .001$. These results along with our previous findings that show that advanced users open Carat less often, yet have better battery life suggest that advanced users have learned to better manage their battery with less help from Carat. Given below is how bugs, hogs, and other application reduction percentages are calculated.

Calculation of Application Use Reductions

We examined the Carat samples and the Carat log reports of the survey respondents. For each user, we obtained the total number of samples u_t they reported to Carat. For the first hog or bug report of each application z for each user u , we split u_t into the samples before the report u_b and after u_a . We took the subset of samples that contained z , before z_{ub} and after z_{ua} the report. Finally, we only considered the samples before the report from the point that z was first run t_{uz1} by u to avoid diluting the ratio: $u_{z1} = u_b | \text{time} \geq t_{uz1}$. Then we compared the ratio of running the application before the report, to after it, and obtained the reduction ratio r :

$$r = 1 - \frac{z_{ua}/u_a}{z_{ub}/u_{z1}}. \quad (1)$$

Note that if the user increased the ratio of running the application, then $r < 0$. We then calculated the averages of decreasing application use for all u in each category, and all z for all the reports that we had obtained from Carat:

$$a = \frac{\sum_u^n \sum_z^m r}{n \times m}. \quad (2)$$

The Respondents' Comments on Carat

We asked the respondents to comment on what they specifically like about Carat, and provide suggestions for improvement. This was an optional part of the survey. Hence only 20.6% of the respondents provided comments. The advanced

Category	% of Hogs/Bugs affected	Hog % reduced	Bug % reduced
Beginners	64.3	36.50	23.48
Advanced users	67.2	46.35	30.12

Table 4. The percentage of Hogs/Bugs that the users reduced the use of, and the reduction percentages, the first time a user gets a report on them.

users provided more suggestions for improvement than beginners, which is natural since they have more experience with Carat.

Most Liked Features of Carat

We received comments about preferable features of Carat from 6.4% of the beginners and 8.6% of the advanced users. Most of the comments were not addressing any specific features, but rather stating general interest in Carat.

The hog and bug reports were positively acknowledged by 12 beginners and 13 advanced users in their comments: “I really like how [Carat] tells you about buggy [applications] and [...] hogs.”

23 advanced users stated that they like the J-Score the most: “The J-Score is a great way for comparing battery life with other devices.” However, only four beginners expressed their interest in J-Score. This is in line with our previous findings about the advanced users being more interested in the J-Score.

14 advanced users and 7 beginners admired non-functional features of Carat, such as reliability, usability, and low battery consumption: “[Carat] just works without being a hog itself.”, “[Carat] works very well and is very simple to use.” Respondents also mentioned using Carat because it does not kill applications by itself but gives control to the user.

Another feature that was mentioned in many comments was the actions tab, and the fact that other applications can be killed directly through Carat. This also gave users information about applications that restart right after killing them: “I have found it useful to see which [applications] [...] are constantly restarted by built-in [...] software.”

Suggestions for Improvement

8.5% of the beginners and 18% of the advanced users made suggestions for improvement. Figure 8 provides a summary of these suggestions. Many of the suggestions were about additional features such as automatic collection of samples, but a significant number of respondents also requested more information about current features. 27 advanced users and five beginners suggested that Carat should send samples automatically or show periodic reminders to open Carat often enough. We also received comments requesting more information about hogs and bugs. Among them we identified three types of problems concerning the actions suggested by Carat: insufficient information about applications reported as hogs/bugs, system applications are suggested for killing, and no solution for applications that reopen immediately after killing. Three beginners and 17 advanced users stated that they would like to have more information about applications that are suggested to be killed, such as what it does, and suggestions for substitute applications. 12 advanced users stated that Carat is

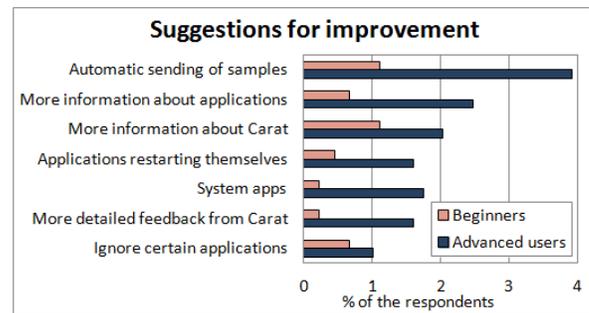


Figure 8. Types of suggestions for improvement

suggesting them to kill system applications: “[...] sometimes [the applications suggested for killing] seem like system [applications] or important services.”. Respondents also stated that they would like Carat to detect applications that reopen immediately after killing, and make alternative suggestions for them. Furthermore, seven advanced users and three beginners stated that sometimes Carat suggests them to kill applications that they use regularly. They prefer a way to hide hogs and bugs that they want to use: “Having a way to ignore some [applications] would be great.”

Five beginners and 14 advanced users commented that they need more information about how Carat works. Beginners stated more directly that they do not understand how Carat works. However, advanced users stated that they misinterpreted some features of Carat, or the feature that the user needed more information about was often specified: “[I] wish I understood what expected improvement means.”

LIMITATIONS

The limitations of self-reporting are well-known. Since the survey respondents were a group of self-opted volunteers among all Carat users, this group might be more interested in features of Carat than the other users. In addition, subjects may, intentionally or not, provide inaccurate or imprecise responses. To address this, we combined Carat log data with user-reported data where possible. However, some aspects cannot be corroborated with Carat log data, such as gender and understanding how Carat works. Since the majority of the respondents were male (89%) the results may not generalize so well to female users and we have no ability to compensate for potential gender biases or incorrect user beliefs on their understanding of how Carat works. For the purpose of our discussion we assume that the gender of respondents is not correlated with the features of interest such as how much their battery life improved.

There could be other external factors such as users’ long term experience with smart phones that could have influenced the battery management skills of users. However, we assume the duration of Carat use is the dominant factor, because previous Carat studies showed that not only does user battery life improve over time, but that this improvement is much stronger for users who receive suggestions from Carat compared with those who do not [15], and also literature [5] suggests that the users form habits with long-term use of energy awareness applications.

Our results track groups, not individual users, and their behavior. In future work, we will conduct longitudinal studies with Carat users and analyze their behavior in more detail.

DISCUSSION AND CONCLUSIONS

We discovered features of Carat that influence user behavior and how behavior changes with long-term use of the application. The findings deepen our knowledge on how to improve community-based battery awareness applications to better support both new and long-term users.

We conducted a survey with existing users of the application, and analyzed their interaction logs from Carat. With the help of these quantitative and qualitative data, we compared the behavior of two types of users, beginners and advanced, to better understand why the latter group enjoys a greater improvement to battery life. Our results revealed that advanced users open Carat less frequently than beginners. However, Carat has considerably changed the mobile device use behavior of advanced users. They have stopped using some applications and replaced them with alternatives, have gained better battery life, charge their devices less frequently, kill reported hogs and bugs more often, and have learned to better manage their battery without the help of Carat. These findings suggest that Carat has changed user behavior while helping users learn to identify applications that drain the battery quickly. Building on these observations, we propose a set of guidelines applicable to the design of battery-awareness applications.

First Guideline: Show Your Work

Carat has succeeded in changing behavior by combining crowdsourcing with explicit instructions that are missing in many similar battery awareness applications [8, 16]. Carat provides explicit information about which applications are draining the battery abnormally quickly through its action list, and the bug and hog reports help the user understand how these applications are affecting the broader community. These are primary features of Carat that enhance user knowledge. Furthermore, information about expected battery life improvement helps users learn how killing an application actually affects the battery life. Our findings indicate that advanced users get into the habit of checking expected battery life improvement before killing applications, and that such features foster learning about mobile battery life and application behavior. In household energy awareness systems, it was recommended to provide feedback to support learning [5]. Our results also suggest that users are not interested in blindly following instructions, but seem to follow Carat suggestions more often when they understand how it works. This understanding helps the user trust the recommendations and possibly learn enough to make similar diagnoses on their own in the future. According to these findings we propose our first guideline: *Expose to the user not just recommendations, but also the reasoning or data behind them.*

Second Guideline: Retain Long-Term Users

If the goal of battery awareness applications is to improve users' knowledge of the device and increase battery life, then prolonged use should result in increasingly better battery life.

In Carat, this effect is amplified by increasingly accurate recommendations as Carat learns more about the user's device. However, our findings also suggest that there is a tendency for long-term users to leave Carat once they have learned to manage their battery without the help of the application. Tailoring features for different types of users has been a challenge in domestic energy awareness research [1, 3]. Community-based mobile battery awareness systems that learn from their users should also be tailored to retain long-term users. The J-Score feature in Carat tries to achieve this retention through community engagement. Advanced Carat users are more interested in the J-Score, and the competitive environment it creates. According to these results, we propose our second guideline: *Tailor community-based battery awareness applications to retain long-term users.*

Third Guideline: Give clear, action-oriented instructions for improving battery life

Providing effective feedback on resource consumption is a key challenge in household energy awareness systems [24]. It is important to give feedback in way that is easy for users to grasp; the instructions should be unambiguous and action-oriented. As shown in Figure 4, the most popular feature of Carat was the "actions" tab. These suggested actions were more popular than the hogs or bugs, even though they simply tell users to "kill" or restart running applications on the hogs and bugs lists. However, in our study we found that the term "killing" an application was misinterpreted by some users, since they feared that killing would result in data loss. The term was chosen to represent permanently closing an application and keeping it closed. Unfortunately, some applications automatically restart when killed, for example Facebook on Android restarts unless "Force Closed" through the Application Manager. Based on these findings we propose our third guideline: *Take into account the audience when formulating feedback to convey precisely what is intended. Provide the user with clear, action-oriented instructions for improving battery life.*

Fourth Guideline: Distinguish System Components

System components pose a problem for Carat, as they are sometimes difficult to distinguish from third-party applications and require different treatment with respect to kill/restart recommendations. Carat maintains a list of system applications in order to mitigate this problem. However, with new versions of mobile operating systems and custom Android versions, maintaining an up-to-date list is a difficult task. Android provides a flag that indicates whether an application is part of the pre-installed image on a device, yet many service providers include applications that can be safely killed such as Facebook and Twitter in the pre-installed applications. This problem can be addressed through crowdsourcing by allowing users to flag suspected system applications. We propose our fourth guideline based on this example: *Distinguish system components from third-party applications when making diagnoses and recommendations.*

The findings presented in this paper provide suggestions for the improvement of mobile battery awareness applications. The guidelines above target community-based mobile battery

awareness applications. Single-device applications can take advantage of all but the second guideline.

REFERENCES

1. Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. *J. Environmental Psych.* 27, 4 (2007), 265–276.
2. Banerjee, N., Rahmati, A., Corner, M. D., Rollins, S., and Zhong, L. Users and batteries: Interactions and adaptive energy management in mobile systems. In *Proc. UbiComp 2007*. Springer, 2007, 217–234.
3. Chetty, M., Tran, D., and Grinter, R. E. Getting to green: understanding resource consumption in the home. In *Proc. UbiComp*, ACM (2008), 242–251.
4. Creus, G., and Kuulusa, M. Optimizing mobile software with built-in power profiling. In *Proc. Mobile Phone Programming*, F. Fitzek and F. Reichert, Eds. Springer Netherlands, 2007, 449–462.
5. Darby, S. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays 486* (2006), 1–21.
6. Datta, S. K., Bonnet, C., and Nikaein, N. Minimizing energy expenditure in smart devices. In *Proc. ICT*, IEEE (2013), 712–717.
7. Ferreira, D., Dey, A. K., and Kostakos, V. Understanding human-smartphone concerns: a study of battery life. In *Proc. Pervasive Computing*. Springer, 2011, 19–33.
8. Ferreira, D., Ferreira, E., Goncalves, J., Kostakos, V., and Dey, A. K. Revisiting human-battery interaction with an interactive battery interface. In *Proc. UbiComp*, ACM (2013).
9. Froehlich, J., Findlater, L., and Landay, J. The design of eco-feedback technology. In *Proc. CHI*, ACM (2010), 1999–2008.
10. Gamberini, L., Spagnolli, A., Corradi, N., Jacucci, G., Tusa, G., Mikkola, T., Zamboni, L., and Hoggan, E. Tailoring feedback to users actions in a persuasive game for household electricity conservation. In *Persuasive Technology. Design for Health and Safety*. Springer, 2012, 100–111.
11. Henryson, J., Håkansson, T., and Pyrko, J. Energy efficiency in buildings through information—swedish perspective. *J. Energy Policy* 28, 3 (2000), 169–180.
12. Liu, X., Shenoy, P., and Corner, M. Chameleon: application level power management with performance isolation. In *Proc. Intl. Conf. on Multimedia*, ACM (2005), 839–848.
13. Midden, C. J., Meter, J. F., Weenig, M. H., and Zieverink, H. J. Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *J. Econ. Psych.* 3, 1 (1983), 65–86.
14. Oliner, A., Iyer, A., Lagerspetz, E., Tarkoma, S., and Stoica, I. Collaborative energy debugging for mobile devices. *Proc. USENIX HotDep* (2012).
15. Oliner, A. J., Iyer, A. P., Stoica, I., Lagerspetz, E., and Tarkoma, S. Carat: Collaborative energy diagnosis for mobile devices. In *Proc. SenSys* (2013).
16. Pathak, A., Hu, Y. C., and Zhang, M. Where is the energy spent inside my app?: fine grained energy accounting on smartphones with eprof. In *Proc. European Conf. on Computer Systems*, ACM (2012), 29–42.
17. Pierce, J., Fan, C., Lomas, D., Marcu, G., and Paulos, E. Some consideration on the (in) effectiveness of residential energy feedback systems. In *Proc. DIS*, ACM (2010), 244–247.
18. Pierce, J., and Paulos, E. Beyond energy monitors: interaction, energy, and emerging energy systems. In *Proc. CHI*, ACM (2012), 665–674.
19. Pierce, J., Schiano, D. J., and Paulos, E. Home, habits, and energy: examining domestic interactions and energy consumption. In *Proc. CHI*, ACM (2010), 1985–1994.
20. Rahmati, A., Qian, A., and Zhong, L. Understanding human-battery interaction on mobile phones. In *Proc. MobileHCI*, ACM (2007), 265–272.
21. Rahmati, A., Tossell, C., Shepard, C., Kortum, P., and Zhong, L. Exploring iphone usage: The influence of socioeconomic differences on smartphone adoption, usage and usability. In *Proc. MobileHCI*, ACM (2012), 11–20.
22. Rahmati, A., and Zhong, L. Human–battery interaction on mobile phones. *J. Pervasive and Mobile Comp.* 5, 5 (2009), 465–477.
23. Riche, Y., Dodge, J., and Metoyer, R. A. Studying always-on electricity feedback in the home. In *Proc. CHI*, ACM (2010), 1995–1998.
24. Rodgers, J., and Bartram, L. Exploring ambient and artistic visualization for residential energy use feedback. *J. IEEE Trans. Visualization and Comp. Graphics* 17, 12 (2011), 2489–2497.
25. Zhang, L., Tiwana, B., Qian, Z., Wang, Z., Dick, R. P., Mao, Z. M., and Yang, L. Accurate online power estimation and automatic battery behavior based power model generation for smartphones. In *Proc. Int. Conf. on Hardware/Software Codesign and System Synthesis*, ACM (2010), 105–114.