Why Would You Do That? Predicting the Uses and Gratifications Behind Smartphone-Usage Behaviors

Alexis Hiniker[1], Shwetak N. Patel[2], Tadayoshi Kohno[2], Julie A. Kientz[1]
Human Centered Design and Engineering[1]
Computer Science and Engineering[2]
University of Washington, Seattle, WA, USA
alexisr@uw.edu, shwetak@cs.washington.edu, yoshi@cs.washington.edu, jkientz@uw.edu

ABSTRACT
While people often use smartphones to achieve specific goals, at other times they use them out of habit or to pass the time. Uses and Gratifications Theory explains that users' motivations for engaging with technology can be divided into instrumental and ritualistic purposes. Instrumental uses of technology are goal-directed and purposeful, while ritualistic uses are habitual and diversionary. In this paper, we provide an empirical account of the nature of instrumental vs. ritualistic use of smartphones based on data collected from 43 Android users over 2 weeks through logging application use and collecting ESM survey data about the purpose of use. We describe the phone-use behaviors users exhibit when seeking instrumental and ritualistic gratifications, and we develop a classification scheme for predicting ritualistic vs. instrumental use with an accuracy of 77% for a general model, increasing to more than 97% with a sliding confidence threshold. We discuss how such a model might be used to improve the experience of smartphone users in application areas such as recommender systems and social media.

Author Keywords
Smartphones; uses and gratifications; machine learning; mobile phones

ACM Classification Keywords
H.m. Information systems: Miscellaneous

INTRODUCTION
Smartphones serve a wide variety of functions, and users can find apps to achieve nearly every imaginable purpose. The reasons that individuals use smartphones has been the subject of a number of prior studies [4,7,23], and past work has shown that there are recurring themes (such as escapism, information-seeking, and communication) in individuals' phone-use motivations [7].

Uses and Gratifications Theory (UGT) explains that technology users are not the passive recipients of the media they consume but are instead active agents seeking out media experiences that meet specific needs. Prior work has established that the uses and gratifications that users seek from digital media can be divided into *instrumental* and *ritualistic* purposes. Instrumental uses of technology are goal-directed and purposeful, while ritualistic uses are habitual and diversionary. A large body of prior work has shown that this dichotomy is relevant to a variety of technologies, including television [18], VCRs [3], social media [16], tablets [9], and feature phones [8]. More recent work has shown that smartphone usage can also be divided into instrumental and ritualistic behaviors [7].

Though existing work has examined the gratifications that users derive from smartphones and established that these purposes cluster into the well-established, higher-level categories of instrumental and ritualistic use [7], the research community has not yet documented the interactions and phone-use behaviors that characterize these categories. The goals of this project were, first, to describe what instrumental and ritualistic smartphone use look like in practice, and, second, to determine whether it is possible to automatically detect instrumental and ritualistic use in real-time.

The ability to characterize and predict instrumental or ritualistic smartphone use has the potential to serve users and designers in a variety of ways. By understanding when users are exploring aimlessly and when users have a specific goal in mind, designers can differentiate product experiences to match these needs. Users who have a specific goal may find suggestions and distractions frustrating, while users who are seeking stimulation may welcome them. Users who wish to better understand their own patterns of behavior might appreciate personal informatics that distinguish their instrumental from their ritualistic behaviors. And designers who wish to understand how users experience their products may benefit from data that describes how their users engage with the application when they have specific intentions and how they engage with the application when they are habitually passing the time.

UbiComp '16, September 12 - 16, 2016, Heidelberg, Germany

ACM 978-1-4503-4461-6/16/09‥$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971762

To explore these questions, we conducted a two-week in-situ study with 43 Android smartphone users. We comprehensively logged participants' application use, and we used the experience sampling method (ESM) to ask users why they were currently using their phone. In this way, we were able to collect a wealth of information about activities that users self-reported as either instrumental or ritualistic.

We found that some apps were used primarily for instrumental or primarily for ritualistic purposes, while others were routinely used for both. With a broad description of a user's recent phone-use – including the series of apps used in the current session, the categories of apps used, the length of the current session, and the time of day – we were able to predict the purpose of phone use with 77% accuracy. Using a probability distribution of confidence, we restricted classification to samples that our model felt 95% certain it could classify correctly, enabling us to extract a targeted 48% of all samples and classify these with 89% accuracy. Restricting the classifier to samples where it was 100% certain it could classify correctly allowed us to include 16% of all samples, and to classify with 97% accuracy.

Our results confirm prior work showing that smartphones are used both instrumentally and ritualistically. We further show that systematic usage patterns are indicative of underlying instrumental or ritualistic motivations. We contribute both descriptive measures characterizing the behaviors that are most likely to reflect instrumental and ritualistic intentions, and we show that it is possible for existing technologies to predict in real-time the kinds of gratifications a user is likely to be seeking.

BACKGROUND AND RELATED WORK

Uses and Gratifications Theory

UGT has a long history as a model for exploring users' interactions with popular media. In the 1940s, scholars began to explore the reasons why individuals chose to listen to radio quiz programs or soap operas [5] or read comic books [25]. This early work framed users' interactions with media as intentional and grounded in users' needs and desires. It explained, for the first time, that users were not passive recipients of the effects of media, but the active member of their partnership with technology, intentionally seeking out the effects that followed from media consumption [21].

In 1972, McQuail and colleagues first began to integrate the diverse taxonomies of uses and gratifications that had been identified across a large body of communication studies scholarship [12]. Alan Rubin further identified that these categories coalesced into what he termed *instrumental* and *ritualistic* purposes: the intentional use of media to achieve a specific aim, and the habitual use of media to pass the time [19].

UGT has since been used to explain a variety of diverse phenomena in human-media interactions. Papacharissi documented that users turn to the Internet in order to avoid face-to-face encounters [13], Leung and colleagues used UGT to identify the reasons why caller ID increases cell phone use [8], and Magsamen-Conrad explained the ways in which older adults' motivations for using tablets differ from those of young people [11]. Despite the many ways in which the technologies we use have changed since UGT was first conceptualized, its constructs and approaches remain relevant and productive [20].

When Rubin initially defined instrumental and ritualistic use, he cautioned that this distinction was not a true dichotomy, and that a variety of contextual factors might cause the same user to engage with the same medium instrumentally at one moment and ritualistically the next [18]. Here, we explore exactly this problem by attempting to learn the behavioral and contextual characteristics that signal instrumental or ritualistic use. By combining a machine-learning approach with UGT's framing, we aim to identify systematic patterns in the diverse ways that users seek out gratifications from their smartphones.

Uses and Gratifications of Smartphone Use

Many studies have probed the gratifications that users seek and obtain from smartphones. Joo and Sang identified a small taxonomy of motivations – including accessing product reviews, keeping up with the news, and relaxation – that describe the reasons college students use smartphones [7]. They further classified these themes through factor analysis under the broader umbrellas of instrumental and ritualistic use. Lina and colleagues identified the reasons why users purchase paid apps [26], including a sense of self-efficacy as a user and approval from peers. Other work has shown that related content satisfies instrumental needs while unrelated content satisfies ritualistic ones when multitasking with smartphones [17]. Bondad-Brown and colleagues documented that the motivations driving the use of earlier technologies extend to our current media landscape [2].

We use this existing work as a foundation that gives us confidence that the divide between ritualistic and instrumental use is meaningful and relevant to smartphones. As prior studies have repeatedly shown that users look to their phones to meet both types of needs, understanding these differing experiences promises to deepen our awareness of the way users relate to their phones. Learning to predict these underlying motivations promises to help designers more effectively provide users with the specific gratifications they are seeking.

Predicting Smartphone Behaviors

A variety of research projects have successfully predicted users' smartphone behaviors or learned about their attitudes by observing their smartphone usage. LiKamWa and colleagues successfully automated the prediction of an individual's average daily mood based on logs of their smartphone usage [10]. Other work predicted when a user was displaying problematic or excessive smartphone usage behaviors with nearly 90% accuracy [22]. A separate research team demonstrated that they could predict whether a user was beginning a new smartphone session or continuing an old one when he or she unlocked the screen [1], while other work used cell tower ID, signal strength, time of day, and day of the week to effectively predict whether or not a user was in close proximity to his or her phone [14]. Pielot and colleagues demonstrated that they could predict a users' level of boredom with 89% accuracy based on the ways in which they had most recently used their phone [15].

These, and many other studies, demonstrate that phone-use behaviors are intricately embedded in daily life, enabling the prediction of a user's moods, habits, and location from their phone activities and vice versa. We build on this work by extending well-established machine-learning approaches to the domain of smartphone gratifications.

METHODS

Participants

We recruited Android smartphone users over the age of 18 through Amazon's Mechanical Turk service to participate in this study. All participants had successfully completed at least 5000 "HITs" (Mechanical Turk tasks) in the past and had a task-approval rating of 99% or above.

Our sample included 43 individuals, with 24 (56%) identifying as female, 16 (37%) identifying as male, 1 reporting non-binary gender, and 2 choosing not disclose gender. The overwhelming majority of participants (88%) reported their race and ethnicity as non-Hispanic White, and approximately half of all participants had an annual household income between $25,000 and $50,000 All participants were living in the United States at the time the data was collected and represented 21 states plus the District of Columbia. Average participant age was 35 (sd = 9.4) and ranged from 22 to 56. Participants had owned a smartphone for an average of 4.6 years (sd = 2.6, min = 3 months, max = 10 years). Comprehensive demographic information is shown in Table 1.

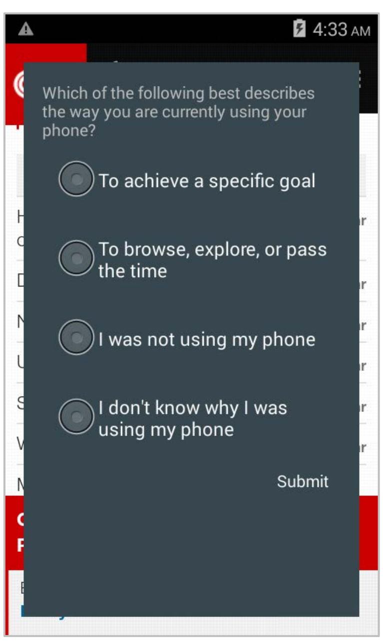

Figure 1: Screenshot of survey, overlaid on active content

Gender	Male (37%), Female (56%), Other (2%), No response (5%)
Age	Mean (sd) = 34.9 (9.4) years Minimum = 22, Maximum = 56
Race and ethnicity	Non-Hispanic White (88%), Black (5%), Hispanic (5%), No response (2%)
Education	High School or Less (14%), Some College (28%), Associate's Degree (9%), Bachelor's Degree (35%), Graduate Degree (14%)
Household income	< $25K (19%), $25-50K (47%), $50-75K (14%), $75-100K (14%), > $100K (7%)
Years of smartphone ownership	Mean (sd) = 4.64 (2.6) years Minimum = 0.25, Maximum = 10

Table 1: Participant demographics

Materials

We implemented an Android application and accompanying background service to deploy on participants' mobile phones. The front-end application displayed a one-question experience sampling survey which asked the user, "Which of the following best describes the way you are currently using your phone?" The user selected from a multiple-choice list of options including: "To achieve a specific goal" (instrumental use) and "To browse, explore, or pass the time" (ritualistic). A screen shot of the ESM survey is shown in Figure 1.

When the survey was displayed, it appeared spontaneously on top of the user's phone without warning. We intentionally overlaid the survey directly on active content (rather than using a push notification) in order to attempt to capture a moment of active use and to avoid leading the user to respond to surveys in between other activities. If the user pressed the back button on his or her phone, the survey was dismissed. If the user pressed the home button, he or she returned to the launcher and the survey was dismissed. Once a survey was submitted or dismissed, it was not possible to retrieve it again.

The background service logged the package name of each app that the user brought to the foreground. It recorded the app name, time of day according to the user's local settings, and the duration of time that the app was active. Both survey data and background logging data were collected using Google Analytics events.

This app also collected ESM data about other topics unrelated to this study. Participants completed an average of 4 unrelated surveys per day, each of which took a few seconds to complete. This data is not analyzed here.

Procedures

Participants were first directed to a screener survey which we used to collect information about their phone use, the reasons they use their phone, and their demographic characteristics.

Altogether, 100 individuals completed this survey. All survey respondents received $1 as a thank-you for their participation.

At the end of the survey, participants were invited to install our Android app and given a link to its page on the Google Play store. Once installed, the application continually collected usage information through its background service. It displayed a survey to the user asking why he or she was using the phone (Figure 1) at random times throughout the day and was displayed approximately once for every hour of phone use. Surveys were intended to be displayed only when the user was actively using his or her phone, though it was not always possible to tell if the user had just put down the phone unless he or she locked the screen.

Of the 100 survey respondents, 53 chose to download the app. Of these 53, 10 did not participate either because the application could not run on their phone or because they chose not to complete the ESM surveys. Participants who installed the app but stopped completing surveys after 2 days or less received $5 as a thank-you for their participation and were told that they were ineligible to continue. All participants who were still participating at the end of the second day continued through the entire duration of the study. Participants received two $25 bonus payments as a thank you for their participation, for total compensation of $50.

Data Analysis

We segmented users' individual streams of app use into discrete sessions, delimited by times they locked the screen, times the screen went dark, and periods of inactivity on a system or launcher window of 30 seconds or more. Because prior work suggests that when a phone is left unused (even for a brief period of time) it is likely that the next usage will represent a different task [1], we used these brief interruptions as new-session indicators.

We identified the package names of the apps that users were actively using when the survey prompt was presented. This generated a list of 161 unique packages across all samples. Using the 42matters API [27], we scraped metadata for each of these package names from the Google Play store. In cases where metadata was not available, we manually searched for the package name to learn more about the app in question.

Using the developer-selected categories for each app as well as the app description, we clustered apps into categories. This was a semi-open coding process, in which we used guidance from pre-defined categories, but reassigned apps to categories when these labels felt inappropriate or too vague (e.g. the "Lifestyle" category did not capture the any of the reasons why a user might use a particular app). After completing a holistic first-pass to identify categories, we performed a directed coding [6] and assigned each app to one predefined category (see Table 2 for all categories).

RESULTS

We collected 1,002 survey responses that were embedded in active phone use and deemed valid. Responses which were sent more than a minute after the survey was presented to the user and responses of "I was not using my phone" were discarded. Participants reported engaging in ritualistic behavior 60% of the time and instrumental behavior 40% of the time.

Apps Used Instrumentally and Ritualistically

We attempted to link our ESM samples to specific apps in several different ways. We first considered the type of app the user was actively using when the survey came up. The most common app categories were Browsing (22.6%), Games (16.7%), Communication (14.6%), and Social Media (14.5%), which together composed more than two-thirds of our sample. We examined the extent to which each of these categories was associated with instrumental versus ritualistic use. A chi-square test revealed that significant differences existed in the distribution of app categories between samples reporting instrumental use and samples reporting ritualistic use ($\chi^2(20) = 244.03$, $p < .001$). Contingency table analysis revealed that most app categories privileged one type of use or the other. Specifically, users were significantly more likely to be using apps in the Browsing, Games, Social Media, News, or Reading categories when seeking ritualistic gratifications. Users were significantly more likely to be using apps in the Communication, Health, Productivity, Coupons and Saving, Utilities, or Maps and Directions when seeking instrumental gratifications.

Despite this divide, apps in some categories were routinely used to seek out both types of gratifications. Though users were more likely to use Browsing applications when seeking ritualistic gratifications, a large minority (36%) of samples

Category	%
Browsing	22.6
Games	16.7
Communication	14.6
Social Media	14.5
Coupons and Saving	6.0
TV and Video	4.9
Music	2.9
News	2.8
Shopping	2.3
Reading	2.2
Photos	1.9
Utilities	1.9
Maps and Directions	1.8
Content Aggregator	1.2
Dating	.9
Health and Fitness	.6
Productivity	.6
Banking	.5
Sports	.4
Education	.3

Table 2: Categories of apps used immediately before a sample was collected

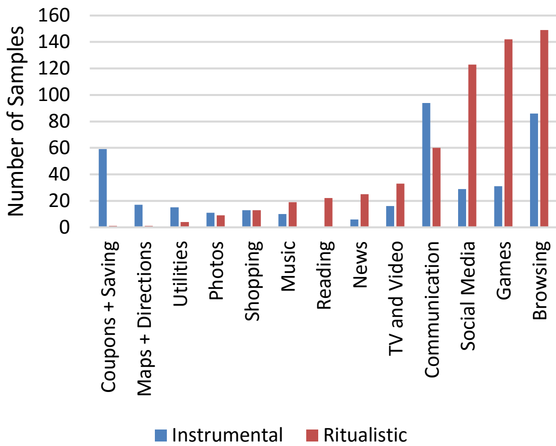

Figure 2: Type of app used immediately before sample was taken as a function of sample type. Only categories linked to at least 2% of samples are shown.

collected while Browsing revealed that participants were engaging in instrumental use. Similarly, while users were more likely to seek instrumental gratifications when using Communication apps, 40% of Communication samples reported ritualistic use. The frequencies with which participants reported instrumental or ritualistic use as a function of the type of app they were using is shown in Figure 2.

We used these same categories to look at the active windows that users brought up immediately after submitting a sample. We saw a similar pattern of results, although using a Browsing app immediately after submitting a sample was not associated with a specific usage type. Finally, we examined the amount of time users spent engaging with apps in each category during the current session up to the point when the sample was collected. Independent-samples *t*-tests revealed that users seeking ritualistic gratifications had spent significantly more time with Browsing, Games, News, Reading, and Social Media apps throughout the current session. Users seeking instrumental gratifications had spent more time with Health and Fitness, Coupons and Saving, Utilities, and Maps and Directions apps throughout the current session. All test statistics are shown in Table 3.

Instrumental and Ritualistic Use throughout the Day

Sample responses were distributed throughout the 24-hour day, with 801 samples (80%) collected between 8 a.m. and 11 p.m. and 201 samples (20%) collected overnight. Because our app detected phone use and displayed the survey approximately once for every hour of use, samples collected at night were collected when the user was already using his or her phone.

We compared the extent to which participants used their phones instrumentally and ritualistically based on the time of day (measured by the minute-of-the-day when the sample was submitted). An independent-samples *t*-test revealed that the average ritualistic use occurred later in the day (mean minute of the day = 899, sd = 383) than the average instrumental use (mean = 840, sd = 343).

Because our minute-of-the-day measure is modular and not continuous, we also bisected the day into morning (8 a.m. to noon), afternoon (noon to 5 p.m.), evening (5 p.m. to 11 p.m.) and night (11 p.m. to 8 a.m.). A chi-square test revealed significant differences in the amount of time participants spent using their phones instrumentally and ritualistically during each of these buckets of time ($\chi^2(3) = 23.07, p < .001$). Post-hoc contingency-table analysis revealed that participants reported significantly more instances of instrumental phone use in the morning than in the evening or

		Mean (sd) seconds	*t*	*df*	*p*
Aggregator	I	2.2 (26)	-1.694	813	.091
	R	7.7 (73)			
Banking	I	1.1 (10)	-.341	1000	.733
	R	1.6 (27)			
Browsing	I	100.9 (251)	-3.276	861	.001**
	R	192.6 (613)			
Com-munication	I	131.9 (496)	1.562	619	.119
	R	88.1 (320)			
Coupons and Saving	I	526.7 (190)	5.538	399	.000**
	R	0.8 (17)			
Dating	I	7.7 (108)	.203	1000	.839
	R	6.6 (61)			
Education	I	0.3 (5)	-1.445	626	.149
	R	2.9 (44)			
Games	I	39.2 (184)	-5.576	741	.000**
	R	194.8 (646)			
Health	I	3.5 (21)	2.562	451	.011*
	R	0.7 (7)			
Maps and Directions	I	15.7 (89)	2.765	493	.006**
	R	2.6 (37)			
Music	I	11.2 (76)	-1.831	722	.068
	R	33.8 (288)			
News	I	6.7 (60)	-2.109	901	.035*
	R	19.7 (132)			
Photos	I	14.1 (129)	.921	1000	.357
	R	8.8 (44)			
Prod-uctivity	I	15.1 (210)	1.406	400	.161
	R	0.2 (6.6)			
Reading	I	3.9 (53)	-3.357	626	.001**
	R	66.2 (451)			
Shopping	I	20.0 (112)	-.256	1000	.798
	R	22.7 (194)			
Social Media	I	56.1 (207)	-2.014	913	.044*
	R	84.2 (230)			
Sports	I	0.2 (4)	-1.544	695	.123
	R	1.2 (15)			
TV and Videos	I	84.4 (445)	-.069	1000	.945
	R	86.3 (382)			
Utility	I	11.8 (77)	1.996	528	.046*
	R	3.5 (37)			

Table 3: Comparison of time in the current session spent on each category of app, up until the point that a sample was collected. I = Instrumental, R = Ritualistic sample.

at night, and significantly more instrumental phone use in the afternoon than in the evening. Participants reported significantly less ritualistic phone use in the morning than in the evening or at night, and significantly less ritualistic phone use in the afternoon than in the evening. All comparisons were relative to expected count.

Thus, instrumental phone use appears to trail off as the day wears on (relative to an individual's baseline level of instrumental use), while ritualistic phone use appears to increase throughout the day (relative to an individual's baseline level of ritualistic use). To visualize this, we constructed a graph plotting the number of samples collected by time of day (see Figure 3). We used a moving average within one hour of each half-hour mark on the graph, such that each half-hour plots the number of samples that we received one hour before and after.

Duration and Rate of Instrumental vs. Ritualistic App Use

Finally, we examined the extent to which users spent time with their phone when seeking instrumental and ritualistic gratifications. To do so, we compared both the duration of the current session up to the point when the sample was collected and the amount of time per window (rate of window-switching) up to the point when the sample was collected. Because both of these variables contained extreme outliers, we used non-parametric statistical tests.

A Mann-Whitney U test revealed that participants spent significantly more time using their phones when seeking ritualistic gratifications (median = 471 seconds, IQR = 800) than when seeking instrumental gratifications (median = 350 seconds, IQR = 806, $U = 2.148$, $p = .032$). Participants spent slightly more time on each individual window when seeking ritualistic gratifications (median = 148 seconds, IQR = 247) than when they were seeking instrumental ones, but this difference was only marginally significant (median = 124 seconds, IQR = 236, $U = 1.875$, $p = .061$).

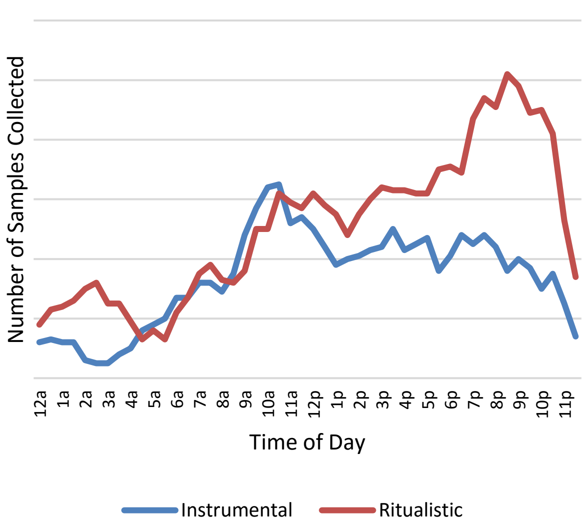

Figure 3: Frequency of instrumental and ritualistic phone use throughout the day

With outliers removed, our results also showed that participants' sessions were significantly longer when they reflected ritualistic use ($U = 2.338$, $p = .019$) and participants spent significantly more time on each window ($U = 2.046$, $p = .041$). This is consistent with the idea that participants intentionally spend time with media to enjoy the experience of using it when seeking ritualistic gratifications but spend the minimal amount of time needed to accomplish a task when seeking instrumental gratifications.

Predicting Instrumental vs. Ritualistic Use

Given that we saw distinctive patterns of behavior emerge for instrumental as compared to ritualistic phone use, we next explored whether we could predict the type of use (instrumental or ritualistic) automatically. To do so, we trained a Decision-Tree-Naïve-Bayes (DTNB) classifier using our set of 1,002 output samples. The DTNB algorithm combines Decision Tree and Naïve Bayes approaches to classification, resulting in area-under-the-curve (AUC) improvements for many datasets over either approach alone [28].

Decision trees and a Bayesian approach have several important advantages for our intent as compared with classifiers such as neural networks or support vector machines. First, they are lightweight and effective predictors that can function easily on a mobile phone. Next, the internal representation of a decision tree is highly human-interpretable and thus can inform application design decisions and feature selection. The features near the root of the tree have high predictive power and can be thought of as the most important features. Therefore, the initial question of what features contribute most to the classification can be addressed. Based on our observation of structure in our participants' phone-use behaviors, this problem also lends itself well to simple probablistic approaches like Naïve-Bayes.

Results with the Full Test Set

We first conducted an exploratory investigation in which we provided our classifier with access to a broad set of features. For each survey response, we extracted a set of phone-use features that we hypothesized could have plausible utility in predicting the corresponding gratification (see Table 4). To do so, we first divided the active session in which a sample was situated into the chain of windows that came before and those which came after. We then calculated the total amount of time that the user had spent in each app category (Browsing, Communication, Social Media, etc.) across all of the windows in the active session that came before the sample was submitted. We identified the name of the app the user was using immediately before the sample was collected, its app category, and the amount of time the user spent with that window. We identified the name, app category, and duration of the app in the current session where the user spent the most time overall (considering only windows the user

Participant ID
Minute of the day
Milliseconds since last session
Milliseconds of use in the past hour
Number of windows viewed in the past hour
Number of sessions in the past hour
Number of windows so far in the current session
Milliseconds per window in the current session
Duration of the current session so far
App name of the window in the current session where the user has spent the most time ("longest window")
App category of longest window
Amount of time spent on longest window
App name of the window the user was viewing when the survey was displayed ("recent window")
App category of recent window
Amount of time spent on recent window
The app category where the user has spent the most time this session ("dominant category")
The percentage of the current session spent in the dominant category
Number of milliseconds spent in each app category this session (20 different features)
App name of the window the user went to after completing the survey ("next window")
App category of next window
Amount of time spent on next window

Table 4: Classifier features

visited before the sample was collected). We also identified the package name, app category, and duration of the window that the user visited immediately after submitting the sample.

Additionally, we extracted the length of the current session up to the point when the sample was collected, the amount of time the participant had spent per window up to the time when the sample was collected, the time of day, and the participant's unique identifier. Finally, we extracted the amount of time the user had spent on his or her phone in the past hour across all sessions, the number of sessions in the past hour, the number of windows in the past hour, and the amount of time between the current session and the one that came before it. Together, these properties yielded 40 features for classification (see Table 4).

We trained our DTNB classifer using all 1,002 output samples with 10-fold cross-validation to avoid overfitting. The classifier selected the participant's unique identifier, the name of the app package that had been used the most in the current session, the category of apps where the user had spent the most time, the amount of time the user had spent with Health and Fitness applications, the amount of time the user had spent with Social Media applications, the amount of time the user had spent with Maps and Directions applications, and the app package for the window that the user visited immediately after submitting the sample, as predictors for its rule-set.

DTNB showed a classification accuracy of 77%. Classifying instrumental samples correctly was more difficult than classifying ritualistic ones, reflected in a lower recall score for our instrumental sample (0.618 vs. 0.87, see Table 5). We compared our classifier against a simple majority-rules classifier (i.e., a classifier which always chooses the most common category); McNemar's chi-square statistic indicated that our classifier demonstrated significantly more accurate performance than this baseline ($p < .001$).

	Precision	Recall	F-Measure
Instrumental	0.76	0.618	0.681
Ritualistic	0.774	0.87	0.819
Weighted Average	0.768	0.769	0.764

Table 5: Performance comparison of a DTNB classifier on instrumental and ritualistic samples

We also examined the performance of our combined approach to that of a Decision Tree or Naïve-Bayes classifier alone. In both cases, the DNTB had superior performance (see Table 6), with gains of approximately 4% over Naïve-Bayes and 7% over ADTree.

	Precision	Recall	F-Measure
ADTree	0.703	0.707	0.696
NaiveBayes	0.729	0.733	0.728
DTNB	0.768	0.769	0.764

Table 6: Performance comparison of a combined DTNB classifier vs. a Decision Tree or Naïve-Bayes approach alone

Given that our descriptive analysis showed that certain app categories were more commonly associated with ritualistic use while others were more commonly associated with instrumental use, we retrained our classifier using app category alone in order to determine the amount of predictive utility this feature carried individually. Using only the app category, our DTNB classifier had an accuracy of 72% (precision = 0.717, recall = 0.72, F-measure = 0.71). Thus, the majority of the discriminatory power of our exploratory feature set could be captured by app category alone, with the full feature set providing an additional 5% accuracy gain.

Finally, we examined whether these predictions changed as a function of the point in the session at which they were asked (for example, near the beginning of a session began or later on). We computed the fraction of the session that had already passed at the point when the sample was collected. An independent samples t-test comparing session-progress in correct vs. incorrect cases showed that there was no difference in accuracy based on the point in the session at which the sample was collected ($t(1000) = -1.034$, $p = .301$).

Results with a Restricted Confidence Threshold

We also considered what it might look like for a system to make use of our classifier's predictions only in instances where the classifier had high confidence that it was correct. Its confidence was trustworthy, such that its accuracy and confidence were roughly equivalent ($r = .953$). Thus, we revisited predictions on our test data, and this time we ignored classifications that were not be made with high confidence. When we set a confidence threshold of 90% (e.g., we required the classifier to be 90% certain of the correctness of its prediction before including the prediction), the classifer achieved 87% accuracy. A confidence threshold of 99% achieved more than 93% accuracy. A confidence threshold of 100% resulted in more than 97% accuracy.

Restricting test samples to those which met the confidence threshold reduced the size of our test set and, in practice, would reduce the number of instances where a real-world system would be permitted to make predictions. To understand the extent to which we were limiting the usefulness of our tool, we examined the number of samples that met the high thresholds that would be necessary in order to achieve high accuracy. By setting a confidence-threshold of 90%, we were limited to 58% of test samples. A confidence-threshold of 99% limited us to 33% of samples. Figure 4 shows the relationship between accuracy and the fraction of samples we were able to include.

Despite the reduced size of our test set, this reduction was spread roughly uniformly across our sample, without privileging specific individuals, types of apps, or time of day. We compared test samples that were included and those that were excluded when the confidence threshold was set to 95% (which led to classification accuracy of 89%). A chi-square test comparing included and excluded samples per participant revealed that the number of samples included with a 95% confidence threshold was significantly lower than expected for 3 of our 43 participants (though one had only 6 total samples). For our other 40 participants, samples were included at a rate that was roughly proportional to the overall distribution of their responses.

We also looked to see if we were systematically excluding data based on other factors that were linked to instrumental or ritualistic phone use. An independent samples *t*-test revealed that there were no differences between the samples we included and the samples we excluded based on amount of time spent browsing ($t(1000) = .812$, $p = .417$), time spent with communication apps ($t(1000) = -.547$, $p = .585$), or time of day ($t(1000) = .087$, $p = .931$). A chi-square test comparing the frequency of samples included or excluded as a function of app category was not significant ($\chi^2(21) = 26.58$, $p = .185$). We also re-ran all of these analyses, splitting data into samples that were included and excluded with a confidence threshold of 99% (which led to classification accuracy of greater than 93%). Again we found that two of our 43 participants had significantly fewer samples included than expected. All other results were non-significant.

DISCUSSION

Characterizing Instrumental and Ritualistic Phone Use

Our results show that several factors have systematic associations with either instrumental or ritualistic smartphone use. The type of app the user is currently using, the types of apps he or she has used most recently, the rate at which the user switches windows, the time of day, and the user's individual habits are all reflective of the gratifications he or she is seeking.

Specifically, our results show that individuals who are seeking ritualistic gratifications are more likely to use social media, play games, scan the news, or read. Participants who are seeking instrumental gratifications are more likely to use utilities, look up savings opportunities, search maps, look up directions, track their health or fitness, or get in touch with others. Users seeking ritualistic gratifications spend more time with each app they use and continue their session for a longer period of time. Users are more likely to seek out ritualistic gratifications and less likely to seek out instrumental gratifications as the day progresses.

However, our results also show that none of these factors alone are perfect predictors of the gratifications the user is seeking, and a deterministic mapping between behavior and its underlying motivation cannot be performed by inspection of individual features. Though Facebook-use was most often associated with ritualistic gratification-seeking, participants regularly reported using Facebook for instrumental purposes as well. The same communication app that a participant used instrumentally at one point became a means of seeking ritualistic gratifications at another.

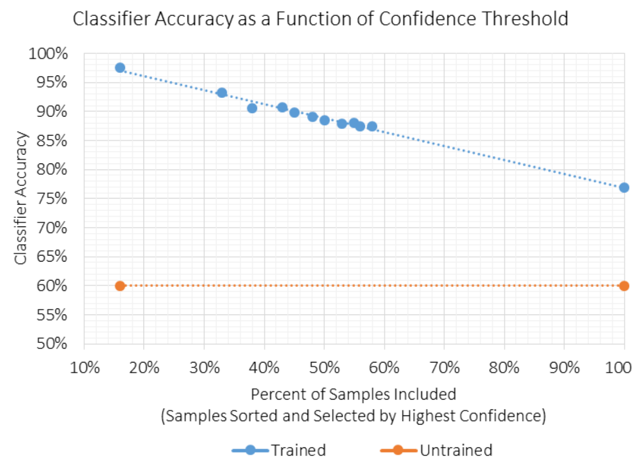

Figure 4: When we included all samples in our test phase, our classifier had an accuracy of 77%. However, as we restricted classification only to samples that the classifier felt it could classify with a certain threshold level of confidence, accuracy improved. In general, the stricter the confidence threshold, the fewer samples we could include in our testing and the more accurate the classifier became. This graph shows the changes in accuracy as we made the confidence threshold more restrictive (thus also reducing sample size). By the time we restricted test samples to those the classifier could classify with 100% confidence, the sample size had shrunk to 16% and accuracy had grown to better than 97%. More usefully, a confidence threshold that restricted us to approximately half our sample achieved an accuracy of approximately 90%.

Rubin cautioned when he first conceived of the instrumental-ritualistic divide that researchers should not attempt to predict the gratifications a user is seeking by considering the medium alone. Consistent with this warning, our results demonstrate that it is inappropriate to brand specific apps as instrumental or ritualistic, rather, it is individual uses of these apps that map back to particular gratifications.

Predicting Instrumental and Ritualistic Smartphone Use

Our results show that the combination of several behavioral features yields predictive power in determing whether a user is actively seeking instrumental or ritualistic gratifications. Knowing the category of the current app was sufficient information for our classifier to achieve 72% accuracy, while a broader range of exploratory features increased accuracy to 77%. App categories were drawn from labels added by the app developer and revised through qualitative coding by the research team. It would be useful to investigate whether other clustering techniques might identify more valuable categories.

Though our simple classifier was only able to achieve an accuracy of 77%, the fact that its confidence estimates were reliable suggests that these features offer further untapped predictive utility (which provide it with the means to know when its guess will be accurate). Combining this classifier with a confidence-threshold heuristic would enable a real-world system to accurately identify a user's current motivation approximately 50% of the time and with approximately 90% accuracy. More importantly, our results suggest that a small set of features describe the structured and differentiated patterns of behavior that users engage in when seeking ritualistic or instrumental gratifications. This suggests that a more sophisticated system with a longer training runway may be able to identifiy users' motivations with even greater precision.

Making Predictions with a Sliding Confidence Threshold

As a mechanism for boosting accuracy, we explored the potential of using the classifier's confidence in its predictions as a filter. We found that we were able to trade the size of our sample for accuracy, suggesting a lever by which developers can improve the predictions of their systems. We saw that the increase in accuracy was dramatic, and that despite the reduction in sample size, we were still able to make predictions across participants and across a diverse set of situations.

One notable limitation of our study is that we began with only 1,002 samples which meant that when we, for example, cut our sample in half (to achieve approximately 90% accuracy), we were left with only 50 test samples distributed across 10 folds, limiting the strength of our claims. Future work remains to evaluate the robustness of the accuracy levels we report here in contexts with additional data.

It is possible that a system trained with a larger dataset or a persistent stream of data may not face this constraint. Developers have the opportunity to optimize not only for accuracy alone, but for the combination of accuracy, confidence, and sample size that leads to the most robust predictions for their particular use case. Future work remains to explore the value of trading off among these factors in predicting systems in other domains.

Differentiated Design for Instrumental and Ritualistic Use

The original vision of ubiquitous computing portrayed a world in which computers are ever-present but recede into the background [24]. Framing users' interactions with technology with UGT suggests that passive, ever-present technology to provide support on an as-needed basis is ideal when users are seeking instrumental gratifications from technology. In instances where users have a specific goal in mind, the most useful role that technology can play is to bring them closer to their goal with as little overhead as possible.

But what of instances in which a user is seeking ritualistic gratifications? In these cases, the user is seeking out experiences without a specific end-point or goal in mind. In these cases the user looks to technology to determine the experience and the user may be more appreciative of direction and suggestion from the interface he or she is using.

Recommender systems might better serve their users with differentiated design for instrumental and ritualistic use. Our participants reported a nearly equal number instances of instrumental and ritualistic motivations when the most recent window they had viewed was a shopping app. A user engaging in ritualistic shopping might appreciate Amazon's recommendation feature that explains "Customers who bought this item also bought…" (see Figure 5) which invites the user to extend his or her experience by browsing items for related scenarios. But a user engaging in instrumental shopping with a concrete goal might find such suggestions frustrating and self-serving.

This divide distinguishes users who want to buy a chair (instrumental) from users who want to shop (ritualistic). Though seeking either of these gratifications might involve a visit to Amazon, understanding the underlying motivation for the visit would allow Amazon to predict the type of recommendations that a user will find most valuable. A user with a specific goal might appreciate knowing what other customers ultimately chose to buy after viewing a particular item, while a user shopping ritualistically might derive more value from recommendations that encourage exploration.

A variety of other technologies might better serve their users by understanding whether they are seeking instrumental or ritualistic gratifications. The popular Chrome extension "Kill the Newsfeed" [29] allows Facebook users to access the site while suppressing the stream of information that they would otherwise see. Presumably, this allows a user to visit the website without being distracted by a bottomless list of enticing alternatives to pursuing their current goal. If Facebook could determine in real-time whether the user was currently seeking instrumental or ritualistic gratification, it

Figure 5: An Amazon feature inviting further shopping that might appeal more to ritualistic gratification-seekers than to instrumental gratification-seekers

might suppress its own newsfeed if and only if the user is attempting to perform a specific, instrumental goal, and then resurrect it when the user returns to the site with ritualistic intentions. This would not only better meet the needs of users visiting the site, it would also eliminate the need for the extreme options that some users currently employ, like suppressing the newsfeed altogether.

By identifying instrumental and ritualistic motivations in real-time, technologies could create different success metrics for serving their users. Technologies supporting instrumental goals might choose to prioritize anticipating questions correctly before the user asks, creating the shortest possible path to success, minimizing the amount of time a user spends on an experience, and keeping the technology and the experience of using as minimal and passive as possible. Technologies supporting ritualistic goals might prioritize presenting a breadth of options, providing an engaging experience that holds the user's interest and inspires him or her to return, and offers stimulation.

Limitations and Future Work

Though our work suggests that combined behavioral features can reveal insights about users' underlying motivations, further work remains to determine which features are most productive and the minimal set of features that are needed to achieve high accuracy. With a larger dataset collected over a longer period of time, it might be possible to create individual models for each user and to evaluate their effectiveness compared to a general-purpose classifier.

Future work also remains to more fully understand what it means for users to be seeking these different gratifications. We did not explore users' interpretation of our question or personal definitions of instrumental or ritualistic use. It is possible that different participants interpreted these prompts differently or had different concepts of what it means to use their phone to pass the time. While we limited the scope of our investigation such that we did not explore users' personal definitions or interpretations of our prompt, it would be valuable to conduct a mixed-methods exploration in the future to better ground our findings in users' mental models of their own behavior and their interpretation of the language we used. Further, it would be useful to determine if users' behaviors cluster into "profiles" of gratification-seeking, i.e., if groups of users seek ritualistic (or instrumental) gratifications in a similar way.

Finally, this work explores the gratifications that users seek without any consideration of the gratifications they actually obtain in practice. It remains unclear whether a user who picks up a phone to pass the time does in fact feel a greater sense of satisfaction after a period of ritualistic browsing or game play. Future work remains to understand the gratifications that users obtain in practice from the design suggestions we present here and to empirically evaluate whether features designed to respond to the gratifications a user is seeking ultimately result in the emotional experience he or she was hoping for.

CONCLUSION

Like many forms of technology that have come before, users leverage their smartphones to achieve both intentional, instrumental purposes and undirected, ritualistic ones. This distinction characterizes the kind of gratifications the user is seeking as well as the types of phone-interaction the user is likely to appreciate. Our results first show that users are more likely to engage in certain behaviors when seeking instrumental gratifications and other behaviors when seeking ritualistic gratifications. The first contribution of this work is an empirical description of each type of smartphone use.

Our results also show that a holistic picture of a user's current behavior can provide predictive insight into the underlying gratifications he or she is seeking. Our relatively naïve, general-purpose classifier for all phone use could detect instrumental vs. ritualistic usage at any moment in time with 77% accuracy. The second contribution of this work is to show that these features can be used together to predict the user's underlying motivation.

Finally, we show that developers creating predictive systems can make trade-offs between accuracy and breadth of applicable use cases to better tailor a system to their specific goals. By restricting our classifier's opportunity to make predictions to moments when it was highly confident in its classifications, we were able to achieve very high accuracy across a range of diverse scenarios.

There are times when users will value experiences that draw them in and capture their attention and imagination. And there are times when users will find these same experiences disruptive and intrusive. Our results provide a path to understanding users' dynamic needs, the motivations behind them, and what our technologies might do in response.

REFERENCES

1. Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, et al. 2016. A Systematic Assessment of Smartphone Usage Gaps. *Proceeding of the thirty-fourth annual CHI conference on Human factors in computing systems (CHI '16)*, ACM Press.

2. Beverly A. Bondad-Brown, Ronald E. Rice, and Katy E. Pearce. 2012. Influences on TV Viewing and Online User-shared Video Use: Demographics, Generations, Contextual Age, Media Use, Motivations, and Audience Activity. *Journal of Broadcasting & Electronic Media* 56, 4, 471–493.
3. A. A. Cohen, M. R. Levey, and K. Golden. 1988. Children's Uses and Gratifications of Home VCRs: Evolution or Revolution. *Communication Research* 15, 6, 772–780.
4. Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. 2010. Diversity in smartphone usage. *Proceedings of the 8th international conference on Mobile systems, applications, and services - MobiSys '10*, ACM Press, 179.
5. Herta Herzog. 1942. Motivations and gratifications of daily serial listeners. *Radio research* 1943.
6. Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9, 1277–88.
7. Jihyuk Joo and Yoonmo Sang. 2013. Exploring Koreans' smartphone usage: An integrated model of the technology acceptance model and uses and gratifications theory. *Computers in Human Behavior* 29, 6, 2512–2518.
8. L. Leung and R. Wei. 2000. More Than Just Talk on the Move: Uses and Gratifications of the Cellular Phone. *Journalism & Mass Communication Quarterly* 77, 2, 308–320.
9. Louis Leung. 2015. Using tablet in solitude for stress reduction: An examination of desire for aloneness, leisure boredom, tablet activities, and location of use. *Computers in Human Behavior* 48, 382–391.
10. Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope. *Proceeding of the 11th annual international conference on Mobile systems, applications, and services (MobiSys '13)*, ACM Press, 389.
11. Kate Magsamen-Conrad, John Dowd, Mohammad Abuljadail, Saud Alsulaiman, and Adnan Shareefi. 2015. Life-Span Differences in the Uses and Gratifications of Tablets: Implications for Older Adults. *Computers in human behavior* 52, 96–106.
12. Denis McQuail, Jay G Blumler, and John R Brown. 1972. The television audience: A revised perspective. *Media studies: A reader* 271, 284.
13. Zizi Papacharissi and Alan M. Rubin. 2010. Predictors of Internet Use. *Journal of Broadcasting & Electronic Media* 44, 2, 175–196.
14. Shwetak N. Patel, Julie A. Kientz, Gillian R. Hayes, Sooraj Bhat, and Gregory D. Abowd. 2006. *UbiComp 2006: Ubiquitous Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg.
15. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce - detecting boredom from mobile phone usage. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*, ACM Press, 825–836.
16. John Raacke and Jennifer Bonds-Raacke. 2008. MySpace and Facebook: applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 11, 2, 169–74.
17. P.L. Patrick Rau (ed.). 2015. *Cross-Cultural Design Applications in Mobile Interaction, Education, Health, Transport and Cultural Heritage*. Springer International Publishing, Cham.
18. Alan M. Rubin. 1984. Ritualized and instrumental television viewing. *Journal of Communication* 34, 3, 67–77.
19. Alan M. Rubin. 1984. Ritualized and Instrumental Television Viewing. *Journal of Communication* 34, 3, 67–77.
20. Thomas E Ruggiero. 2009. Mass Communication and Society Uses and Gratifications Theory in the 21st Century. June 2014, 37–41.
21. Wilbur Schramm, Jack Lyle, and Edwin B Parker. 1961. *Television in the lives of our children*. Stanford University Press.
22. Choonsung Shin and Anind K. Dey. 2013. Automatically detecting problematic use of smartphones. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*, ACM Press, 335.
23. Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. 2008. A diary study of mobile information needs. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems (CHI '08)*, ACM Press, 433.
24. Mark Weiser. 1991. The Computer for the 21st Century. *Scientific American* 265, 3, 94–104.
25. Katherine M. Wolf and Marjorie Fiske. 1949. The children talk about comics.
26. What Makes Users Buy Paid Smartphone Applications? Examining App, Personal, and Social Influences - ProQuest. Retrieved April 1, 2016 from http://search.proquest.com.offcampus.lib.washington.edu/openview/fbabbf88670f9a9e1f1e7b8bbf435f75/1?pq-origsite=gscholar

27. 42 Matters. Retrieved January 4, 2016 from https://42matters.com/
28. Combining Naive Bayes and Decision Tables. Retrieved April 1, 2016 from http://www.aaai.org/Papers/FLAIRS/2008/FLAIRS08-076.pdf
29. Kill the Newsfeed. Retrieved January 4, 2016 from https://chrome.google.com/webstore/detail/kill-news-feed/hjobfcedfgohjkaieocljfcppjbkglfd?hl=en