Mobile Microwaiting Moments: The Role of Context in Receptivity to Content While on the Go

Ellen Isaacs, Nicholas Yee, Diane J. Schiano, Nathan Good, Nicolas Ducheneaut & Victoria Bellotti
Palo Alto Research Center (PARC)
3333 Coyote Hill Road, Palo Alto, CA 94304
eisaacs, nyee, dschiano, ngood, nicolas, bellotti@parc.com

ABSTRACT

With the growing interest in delivering content to mobile phones, there is a need to understand the contexts in which people are receptive to receiving content on the go. We studied receptivity to mobile content and found that location, activity, and other context factors predicted receptivity to a limited degree, and that the contexts in which people were receptive to pushed content were different from those in which people requested content. People frequently requested content during "microwaiting" moments – brief intervals of idle time *within* other ongoing activities, such as brewing coffee, working out, or getting gas. People also wanted control in determining when to receive content. Our findings suggest that it is difficult to use context information to find the perfect time to push information to people, but it may be helpful in improving the quality of content delivered when it is requested.

Author Keywords

Mobile computing, mobile content delivery, context-aware computing, location-based services, interruptions, user studies, experience sampling, diary studies.

ACM Classification Keywords

H5.m. Information interfaces & presentation: Miscellaneous

INTRODUCTION

There is great interest in collecting contextual data from mobile phones and using it to provide relevant information and services to people on the go. Many research projects aim to use mobile context data to provide such information as the movement of people and resources within a city, nearby social activity, and even appropriate use of foreign language terms [6, 14, 27, 28, 32]. More commercially oriented location-based services send information and advertisements to users based on their location [12, 29, 30]. Some of these services are designed to push information to people and others make it available for people to request, either from a mobile phone application or a web site.

One premise behind this work is that, if given enough information about people's context, we can predict what type of information they will be interested in, either right away or in the near future. Some studies are beginning to collect data to support this effort [8, 34] by asking people to keep track of the information they would have liked in different contexts and then characterizing the factors that affected their content needs.

A second assumption behind services notifying people of useful content is that, once we know what content people want, we will be able to deliver it at appropriate times. Based on research on interruptibility in the workplace [13, 21], however, there is reason to believe that this is not a simple problem. Even if content is useful, people may be annoyed by it if it interrupts them at an inappropriate moment. Or they may ignore an alert if it comes at a bad time and only discover the information later when it's too late to make use of it. This study is aimed at addressing this issue. If sensors can provide us with a rich understanding of people's context, how well can we determine appropriate times to deliver content to them? Put another way, how does receptivity to mobile content differ in various contexts, and how well can we predict this receptivity?

One way to address receptivity is to send content to people's mobile phones in different contexts and get a measure of their receptivity each time, hoping that we can see patterns of high and low receptivity across the contexts. Another approach is to let people request content whenever they wish and look at the contexts in which they do so. In this study, we take both approaches. In doing so, we attempt to tackle several research questions. First, which specific factors affect people's receptivity to content while on the go? Second, how well can we predict receptivity using a machine learning approach, and which factors or sensors are the best predictors of receptivity? Third, are the contexts in which people report being more receptive to *receiving* content equivalent to the contexts in which people are interested in *requesting* content? That is, would a model of receptivity to "pushed" content accurately predict the contexts in which people would "pull" it?

Before going further, we need to clarify what we mean by context. Our goal is an applied one in that we want to determine how well the information that can be collected from sensors can be used to determine receptivity to mobile content. As such, we start with what Dourish refers to as a *positivist* approach [11] and define context as location, time, number of people, noise level, lighting, and indoor/outdoor setting, i.e., factors that sensors can determine. Recognizing that one's activity is also likely to be a critical component of receptivity, and that some research aims to determine activity based on sensed factors [18, 24], we also include "activity" as part of our working definition of context. In this way, we attempt to incorporate the spirit of the *interactional* approach to context as described by Dourish [11]. In this view, context arises dynamically out of activities combined with locations, settings, and objects, and depends on the actor's attitudes about what is relevant at any given time. We will return to this view of context in our discussion.

RELATED WORK

Researchers are viewing the network of mobile phones around the world as a "global mobile sensing device" [6] that can detect people's locations, movements, communication activity, and physical environment while also enabling users to contribute text and images of their environments [6, 14, 28]. The data are then presented to show the pattern of activity in an area, such as on city streets, ski slopes, bike paths, or at local attractions [6, 32]. The idea is that people will retrieve the content, either from their mobile phone or computer, to get an overview of various types of human activity patterns, and they will be notified on their phones when information of interest is available. For example, eNcentive [30] notifies people of promotions or discounts on their mobile phones based on their location and preferences. MicroBlog [14] allows Web users to post location-based questions that are pushed to mobile users in those environments in the hopes that some will respond in short order. In these studies, little is mentioned about how to integrate these alerts into people's activities non-disruptively. Indeed, a preliminary user study of MicroBlog found that queries sometimes arrived too late or at a bad time [14].

Meanwhile, a great deal of research has examined interruptibility at work with the aim of reducing the well-documented impact of interruptions on task completion time and affective state [3, 4, 9]. One approach has been to systematically compare interruption costs at different task phases and transition points [5, 12-15]. Studies have found that interruptions during natural breakpoints [20] or task phases with lower mental workloads [19] result in less disruption. Another approach has been to develop personal interruption cost models based on training data where users are asked to indicate their availability [2, 17, 23]. These machine learning approaches have shown that predictive user models can be created to infer a user's interruptibility.

Fewer studies have explored interruptibility in other contexts. Those done in the home have consistently identified location as a significant predictor [26, 35, 36, 38]. For example, participants rated interruptions in the kitchen as more acceptable than in bedrooms [26, 36]. Time [36] and message urgency [38] have also been found to affect interruptibility. But the results have been mixed. While some studies have found activity to be indicative of interruptibility [26, 35], others suggest that activity may be too coarse a factor to be predictive. For example, participants in one study rated watching TV as both a high and low engagement activity depending on their goals [38].

A study that examined information needs on the go found that people most often wanted to get answers to "trivia questions," i.e. specific questions that arose during a conversation, and that location was the most widely cited prompt for information needs [34]. And surprisingly, survey studies of reactions to mobile advertising found that participants rated the entertainment value of a targeted advertisement as more important than its credibility or informativeness [33, 37].

From a methodological standpoint, research on interruptibility at home and at work has typically gathered data using experience sampling methods [e.g., 17], i.e. interrupting people at random times and asking about their activity and state of mind. In contrast, studies of information needs usually employ diary methods that ask people to note instances when certain events occurred naturally [e.g., 8, 34]. For example, Sohn et al. [34] asked people to write a short note on their phones any time they had an information need while mobile. Brown et al. [5] asked participants to take photographs of situations in which they wanted information captured for later use. In our study, we make use of both methods to study the two main modes of information delivery on mobile devices — "push" and "pull." Experience sampling lets us examine how users react to receiving information at random times, while a diary method allows us to explore those times when people are actively interested in seeking information.

METHOD

Participants

Thirty-two people (16 women and 16 men) from the San Francisco Bay Area participated in the study. They varied in age (18 to 58 yrs, mean = 35), in education level (from some high school to advanced degrees), and were engaged in a wide range of occupations (e.g., teacher, gardener, civil engineer, marketing director, stay-at-home parent). Participants were randomly assigned to the Push and Pull condition, balancing for gender, age, and education level.

Procedure

The study began with a 90-minute session that included training, an initial survey, and an interview, during which participants told us about their interests, activities, and information needs, particularly while on the go. Over the next 14 days, participants received text messages on their cell phones. Since we were focusing on context and not content, we sent people a single type of content that was meant to be reasonably entertaining and quick to read – namely short pieces of trivia, similar to what Sohn [34] found to be most commonly desired. These trivia "snippets" were collected from trivia-related websites. A typical example is, *"There is an ant in Brazil that has a gland which causes the ant to explode like a bomb, spraying a* *sticky toxic goo on everything nearby."* The snippets were no more than 160 characters long, the SMS standard.

An SMS server was set up to deliver content to participants' phones. For participants in the Push condition, it automatically delivered six trivia snippets per day at random times within at least a 10-hour range specified by the participants, with a minimum gap of one hour between snippets. Participants were instructed to read the messages when they arrived and reply to them with a rating of their receptivity on a scale of 1-7 (1=not at all receptive, 7=very receptive), along with a short comment to remind them of their context at the time, e.g. "coffee break." Participants were asked to respond as soon as possible, and if they did not respond before receiving a new message, no rating was recorded for the earlier message. Participants were told to rate their "gut reaction" to receiving the message, taking into account anything they considered relevant. Participants could also include in their reply a request for more snippets.

Those in the Pull condition were told to send a text message to the SMS server requesting a trivia snippet any time they were interested in getting some information. To give them an incentive, the trivia was sometimes replaced by "money snippets," which arrived at random times on a variable ratio schedule and promised the participant a random amount between 50 cents and $3.50, with a maximum available per day of about $14. When the snippet arrived, the participant could request another one as many times as they wanted. They too were asked to append their requests with a few words to remind them of their context.

Participants in both conditions were required to fill out a survey as soon as possible but no longer than a day later. The survey asked a series of detailed questions about their context at the time they received the snippets, including their location, the number of people they were with (social context), their activity immediately prior to receiving the snippet, their use of media, and characteristics of their environment such as amount of light, noise levels, and indoor/outdoor setting (a simple way of simulating plausible sensor readings). Participants were also required to rate how interesting they found each snippet. Most questions requested multiple-choice responses with a text field for clarification, which sometimes was required.

In practice people generally filled out the survey at the end of the day or early the next day. If they did not fill it out within a day, that day's questions were no longer available and participants lost credit for that day's participation. The survey displayed the content of the snippets along with the time each was received and the participant's comment reminding them of their context at the time. In the Push condition participants filled out the survey for each of the six snippets they received, even if they hadn't responded with a receptivity rating. In the Pull condition, if participants had pulled more than six times or "sessions" per day, only six of them were chosen at random to be presented in the survey. A session was defined as a series of snippet requests each within five minutes of one another. If they pulled fewer than six times per day, they answered questions about all their snippet sessions.

At the end of the study, we interviewed participants again for an hour regarding their experiences during the study. All participants were then paid based on the number of daily surveys they completed. Push participants received a fixed amount per day up to a total of $200, and Pull participants received the amount of money they had been awarded in the money snippets each day, up to a possible total of $200 if they completed all their daily surveys.

RESULTS

The 32 participants received 10,055 snippets over the two-week period, 1,933 for Push participants (8.6 per person per day) and 8,122 for Pull participants (36.4 per person per day). Pull participants requested snippets over a total of 1,636 sessions (i.e. a series of requests with a gap of no more than five minutes between requests), with an average of 5.0 snippet requests per session and 7.3 sessions per day. Push participants, who were allowed to pull for snippets as well, received snippets over 1,482 sessions, averaging 1.3 snippets per session. They responded with a receptivity rating for 80.0% of the messages they received. The daily survey sampled 2,395 of those Push and Pull sessions. In 30 cases a participant said she or he didn't remember or didn't get the snippet, leaving 2,365 sessions for detailed analysis.

Since the Pull participants received more than four times as many snippets as Push participants, we investigated the effect of the monetary incentive. Pull participants were more likely to pull for another message after receiving a money snippet vs. a trivia snippet ($F(1,15)=9.35$, $p<.01$), but the amount of money received did not affect whether they pulled again ($F(3,15)=.52$, ns. Further, we found that 68.7% of the time, Pull participants did not reach the maximum payout. In fact, 25% of the time they earned no more than $2.31 and 50% of the time they earned no more than $8.48. In cases when they did earn the full $14, they continued to pull for an average of 28.8 more snippets (SD=42.2). So although the money was a good motivator, it was not the only factor driving people to request snippets.

In the interviews, several in the Pull condition said that they initially requested snippets for the money, but then found themselves getting interested in the information. One person explained, *"At first I was interested partially with the trivia, but mostly with money. And then once it became clear that I was going to get the maximum, I suddenly started pulling for the trivia and I kept on telling it to people. And so my interest had totally changed."* Another also said after pulling initially for the money, *"then I got one that said the venom from a black widow is more poisonous than a rattlesnake. It … got kind of interesting. So I started getting more and more of them. … Then I'd call my brother and say 'Hey, did you know this?' You know, that made me look intelligent."*

To analyze the data from the daily survey, we reviewed the participants' text comments along with their multiple choice responses and, for location, activity, and media use, placed the responses into mutually exclusive categories derived largely bottom-up, following a grounded theory approach [7]. For location and activity, we developed a two-level hierarchy scheme, with fewer *general* categories and more *specific* categories. For location, we identified 5 general location types (shown in Figure 1) and 19 specific ones (shown in Table 1). For activity, we identified 9 general activities (informed by the U.S. Bureau of Labor Statistics' American Time Use Survey [1]), shown in Figure 2, and 22 specific ones, shown in Table 2. For simplicity, when people were doing multiple activities we chose the one judged to be the primary activity, except when multiple activities co-occurred often enough to be seen as a unique activity.

In the case of location and media use, one author did the coding, as the responses were relatively unambiguous. The activity responses were much less straightforward, so two of the authors independently coded them and achieved 79.1% level of agreement. Differences were resolved until 100% agreement was reached for further analysis.

Overview of Receptivity Scores

We begin our discussion of results by looking at the receptivity scores from the Push condition. Not surprisingly, receptivity was affected by the message content. Across all contexts, people rated themselves more receptive when receiving snippets they rated more interesting ($r = .29$, $p < .001$).

On the whole, most of the Push participants also said they liked the trivia snippets, and although the timing wasn't always good, they still liked reading them. A typical comment was: "*I kept looking forward to it. There were some pretty interesting pieces of information, so I thought it was pretty cool.*" Another explained, *"there were ones that it was just, like, a really bad time for me to receive a message, but I still liked seeing the message."* Still, that person said that by the end of the study, *"I just like resented my phone. I … felt like a slave to my phone."*

The overall receptivity score for the Push condition was 4.1 (SD 2.1), just above neutral, even though the snippets arrived at random times. Another study that interrupted people but did not provide any content received negative receptivity scores on average when interrupting at random, improving to neutral only when they were timed to arrive at transition points [9]. So being interrupted with trivia appears to be better than just being interrupted, and overall garners a neutral reaction.

The descriptive statistics give an initial picture of how different aspects of context affected people's receptivity to pushed content. We look specifically at the Push participants' receptivity scores based on location, activity, social context, use of media, noise level, light level, and indoor/outdoor setting. We then go on to see how well all those factors *predicted* receptivity to pushed content.

Location

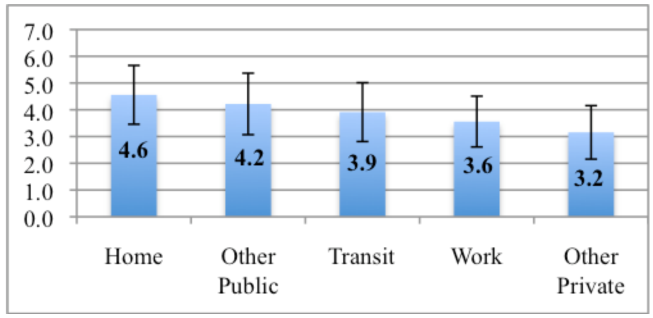

Figure 1. Average Receptivity Scores by Location (1=Not at all receptive, 7=Very receptive)

Figure 1 shows the Push participants' receptivity scores across *general* locations. They were most receptive to receiving trivia snippets while at home and in public settings and least receptive when in other private homes, usually a friend's. Overall, though, the range in receptivity scores across locations was modest ($F[5, 1116] = 10.73$, $p < .001$, $\eta^2 = .05$), relative to the effect sizes generally needed to predict behavior.

Location	N	Mea	SD
Library	15	6.3	1.9
Mass transit	25	5.8	2.2
Gym	12	5.1	2.5
Transit station	11	4.7	2.3
Store	44	4.6	2.2
Home	502	4.6	2.2
Service outlet (bank, salon, gas station,	11	4.6	2.2
Restaurant	49	4.4	2.0
Street	15	4.4	2.3
Work	205	3.5	1.9
Car	86	3.4	1.9
Other public building	4	3.2	1.3
Other private home	75	3.1	2.1
Professional's office	9	3.0	1.7
Theater	16	2.8	2.3
Other outdoor	6	2.7	2.3
Park	10	2.6	1.8
School	4	2.5	2.4
Other indoor	11	2.2	2.0

Table 1. Receptivity Scores for Specific Locations

The 19 *specific* locations showed a wider range of receptivity ($F[28,1109] = 26.02$, $p < .001$, $\eta^2 = .13$). As shown in Table 1, people were most receptive in the library and on trains and buses and least receptive at recreational facilities, such as bowling alleys, parks, pools, arenas, and theaters. It appears that to distinguish receptivity across locations one must to be able to discriminate between fairly specific types of settings.

Activity

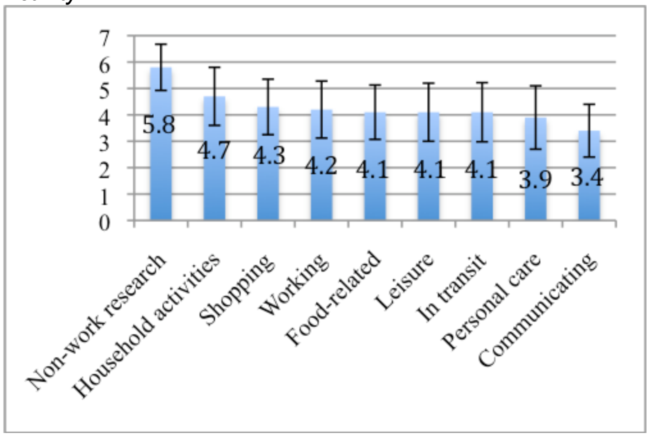

Figure 2. Average receptivity scores by general activity (1=Not at all receptive, 7=Very receptive)

Receptivity across general activities also showed relatively modest differences except at the ends of the spectrum ($F[23, 1132] = 5.30$, $p < .001$, $\eta^2 = .10$), see Figure 2. Participants were most receptive when researching information for non-work purposes, usually on the computer (5.8), and least receptive communicating (3.4), but their receptivity varied little for most other activities, such as shopping, eating or preparing food, in transit (including walking), or engaged in personal care (e.g. grooming, getting dressed).

Activity	N	Mea	SD
Eating & reading	4	5.8	2.5
Reading	35	5.5	2.2
In transition	48	4.9	2.3
Passenger	11	4.8	1.7
Using computer	192	4.8	2.0
Relaxing	25	4.6	2.3
Watching/Listening	113	4.5	2.1
Grooming	51	4.5	2.2
Shopping	31	4.5	2.1
Walking	17	4.4	1.8
Waiting	39	4.2	1.8
Chores	93	4.1	2.2
Cooking	38	4.0	2.0
Eating	76	3.9	2.2
Exercising	11	3.6	2.9
Sleeping	54	3.4	2.6
Conversing/Interacti	122	3.3	2.1
Driving	77	3.3	2.0
Eating & watching	12	3.0	1.4
Meeting	31	2.8	2.2
Creating	12	2.7	2.2
Waking up	9	2.4	1.5

Table 2. Receptivity Scores for Specific Activities

Once again, we saw a wider range of receptivity across the *specific* types of activities, see Table 2. Participants were most receptive to pushed content when reading printed material, transitioning between activities, being driven in a car, using the computer and relaxing. They were least receptive when waking up, creating (e.g. knitting, photographing), or in meetings.

Social context, media use, and sensor data

In terms of social context, participants reported being more receptive when they were with three or more people (4.4, SD=2.2) or alone (4.3, SD=2.3), and less so when they were with one other person (3.4, SD=2.1) or two others (4.0, SD=2.2). When looking at the media or material people were interacting with, we found that people were most receptive when interacting with printed material (5.3, SD=2.3) or the computer (4.8, SD=2.0). They were least receptive when engaged in some form of in-person interaction, such as face-to-face conversations, shopping in stores, and attending live events (3.5, SD=2.2), and when they were already using their phones (3.5, SD=2.1).

We also explored how much receptivity was associated with the simulated sensor measures in our survey, specifically noise level, lighting, and indoor/outdoor setting. Participants reported significantly higher receptivity when noise levels were lower ($r = -.06$, $p = .04$), although the effect was quite small. Participants reported higher receptivity with brighter lighting levels ($r = .13$, $p < .001$). Their receptivity did not differ significantly when they were indoors vs. outdoors ($p = .25$). Finally, we looked at differences across time of day, day of the week, and weekdays vs. weekends and found no significant differences in receptivity.

While we did find differences in receptivity across different contexts, the key finding from these descriptive statistics is that people's receptivity to pushed content varied only mildly across most of the context attributes examined. We saw a wider range of receptivity only when we broke down the context attributes into many specific categories, such as for location and activity. This suggests that systems attempting to deliver content to people at times when they are most receptive need to have a fine-grained model of context. Although the specific numbers are bound to vary across people and situations, the data presented here may give some guidance regarding the level of specificity required and general trends in receptivity.

Predicting Receptivity

So far we have looked at only one context factor at a time. What we'd like to know is how well we can predict receptivity to pushed content if we combine all the variables. To answer this question we used a statistical method called classification trees, a standard machine learning technique. First, we converted receptivity scores to binary values, with receptivity scores of 1 and 2 considered *Low,* and receptivity scores of 6 and 7 considered *High*. Our data set happened to have nearly the same number of low and high scores (356 and 357 respectively), so chance performance in predicting high receptivity would be 50%. To determine classification accuracy (CA), we used five-fold cross-validation. The model was created based on 80% of the data and then tested on the remaining 20%. This procedure was performed five times until all data had been rotated through as test data. The overall CA was the average of individual CAs across the folds.

Table 3 shows the CAs of each variable. Consistent with our exploratory analysis of the descriptive statistics, we found that specific location was best able to predict whether participants' receptivity was high or low. It did so at 66% accuracy, somewhat above chance. Specific activity and interest in the content were slightly less predictive. Number of people and time performed barely above chance.

Category	CA
Specific location	66%
Specific activity	61%
Interest in content	66%
Simple Sensors	59%
Media use	58%
Number of people	54%
Time	53%

Table 3. Classification accuracy of context variables in predicting receptivity scores. Simple sensors includes noise, lighting, and indoor/outdoor setting.

With all the context attributes combined (using five-fold cross-validation), accuracy in predicting receptivity to pushed content improved to 68%, still only a little better than specific location alone. In fact, this level of accuracy was achieved with just three variables: specific location, specific activity, and content interest rating. That is, if we know the participants' specific location type, specific activity type, and how interesting they find the content, we can predict whether their receptivity will be high or low with 68% accuracy, 18% better than chance. Knowing all the other aspects of participants' context did not improve the model's ability to predict receptivity.

Systems that notify people of content won't be able to determine in advance the person's level of interest in the content, of course. Without knowing interest level, the best the model could do in predicting receptivity was 66% – better than chance but not especially impressive.

Using Receptivity to Pushed Content to Predict Pulling for Content

So far we have explored only Push participants' receptivity scores. Looking at the contexts in which the Pull participants requested snippets gives us another view into receptivity. We start by asking whether the two are similar. That is, how well do Push participants' receptivity scores predict when people will pull for content? To answer this, we assume that people pulled only when they were highly receptive and then test the machine learning model based on Push participants to see how often it predicts that Pull receptivity is high. We found that the model predicted the Pull data with a CA of only 51%, essentially chance.

Apparently, the contexts in which people choose to request content are different from those in which people are receptive to pushed content. The following sections explore how those contexts differ.

Requesting Snippets

Since the Pull participants' data did not have receptivity scores, we instead looked at the frequency with which they requested snippets in different contexts. People were not in all contexts equal amounts of time, however, so we need to compare those Pull frequencies with a baseline that indicates how often they were in those situations. Since the Push participants received snippets at random times, we use their frequency data as this baseline measure. By comparing the two, we can see when people requested snippets more or less often than expected by chance.

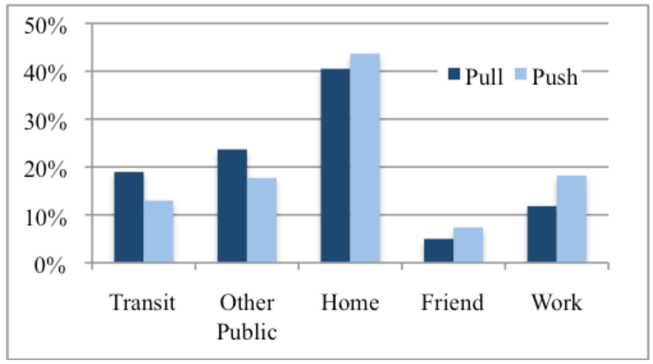

Figure 3. Percentage of snippets requested (Pull) and received (Push) in general locations.

We found relatively small differences in the frequency with which Pull and Push participants received snippets in terms of location ($\chi^2[5] = 45.22$, $p < .001$, $r = .14$), media use ($\chi^2[9] = 25.45$, $p = .003$, $r = .10$), and social context ($\chi^2[5] = 13.06$, $p = .02$, $r = .07$). Figure 3, which shows the differences in snippet frequencies across general locations, represents a typical pattern. Pull participants were a little more likely than expected to request snippets in public settings (especially restaurants and parks) and while in transit, and a little less likely to do so from work, but the differences, while statistically significant, were quite small.

Similarly, participants' general activity also had only a small effect on whether they were inclined to request snippets. However, the 22 specific activities did reveal large differences in the rate at which people requested snippets ($\chi^2[23] = 348.64$, $p < .001$, $r = .39$), as shown in Figure 4. Compared with the base rate, people were 4.6 times more likely to request snippets while Waiting. They were also much more likely to request snippets while Eating & Reading (4.3x), riding in a car as a Passenger (3.2x), Walking (3.2x), Waking up (1.9x), and In Transition between activities (1.9x). On the other hand, people were much *less* likely to ask for snippets while Sleeping (of course), Grooming (including showering and using the bathroom) (8.0x), Eating alone while Watching TV (3.7x), doing Chores (3.4x), Shopping (3.1x), Exercising (2.8x), and in a Meeting (1.9x).

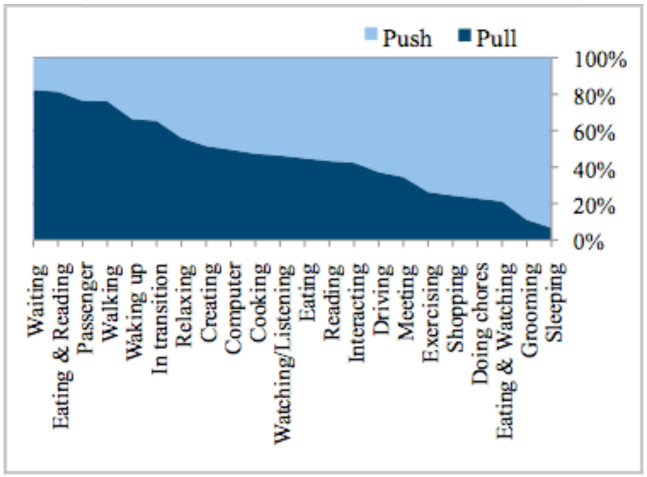

Figure 4. Ratio of frequency of snippets pulled vs. pushed during specific activities.

What surprised us were some of the activities during which these differences were relatively small. For example, we would expect people to be much less likely to request snippets when they're Interacting with others (usually in conversation), while Cooking, and certainly while Driving, and yet those activities came out roughly comparable to the Push base rate. They pulled while Cooking almost exactly as often as we would expect by chance (1.1 times less often), and while Interacting or Driving only slightly less often (1.4x and 1.7x, respectively).

To understand this pattern we looked more closely at the comments offered in the daily survey explaining when and why people requested snippets. The biggest clue came from the times when people pulled while Waiting, which happened nearly 5 times more often than the base rate. In some of these cases, waiting was the main activity, as when waiting for transit *("Got off train, waiting for BART")*, for a person *("waiting for doctor to get back")*, for service *("I just ordered my food and was waiting for it")*, or in line *("I was waiting in line to be let into the safe deposit box")*.

But many of them were cases of what we call *microwaiting moments*, or short periods of idle time in the middle of other ongoing activities, such as pumping gas *("sitting in the car, waiting for the car's gas tank to be filled up")* or working on the computer *("the files were taking a while to upload.")*

It turned out that these microwaiting moments also appeared during many other activities, such as exercising *("I had free time in between workout sets"),* preparing food *("I was grinding coffee beans, then waiting for water to heat")*, playing games (*"there was a break in the [video] game because my friend needed to fool around with something solo,"),* and shopping*, ("I was waiting for the salesperson to ask some questions").* Microwaiting moments also happened when people were driving, as when one person was *"waiting for a parking space, the other car was taking a while to leave,"* or when *"a train passed on my route."* People also found microwaiting moments during conversations, e.g., *"there was a break in the conversation,"* and *"my husband had to put me on hold.*"

It appears that throughout the day, people found many idle periods, maybe 30 seconds to a few minutes long, that were ideal times to request snippets. One person explained in the interview, *"most likely I was doing something else, either sitting at a red light or as I said I was on a conference call once at my office and requested it. So I'm quite a multi-tasker. So sometimes it would just give me something to do while I was kind of on hold with another activity."*

Sometimes people used the snippets as a way to take a short break from an activity, e.g. *"I wanted a quick break from the book"* and *"packing – needed a break."* In a few interesting cases, people even used the snippets to distract them from an unpleasant ongoing activity, for example, *"It was cold outside, trying to get my mind off how cold I was,"* or *"trying to block out the background noise of people"* on a bus. One woman even requested snippets during the fight scene in a movie because she didn't want to watch it.

This finding is consistent with that of Naaman [25], who also reported evidence of people using short moments to do mobile phone activities, in that case browsing photos. They found that 50% of photo browsing sessions were shorter than two minutes, and 12.5% lasted under 30 seconds. On average, the Pull sessions in our study lasted 3:46 minutes, during which people retrieved 6.1 snippets. Our finding is also consistent with the results of workplace interruptibility studies, which show that interruptions were less disruptive during natural breakpoints and during tasks requiring lower mental engagement [20]. It appears that those convenient break points may be more common and more varied when people are mobile.

In summary, the machine learning model showed that receptivity to pushed content did not predict when people requested content. When we pushed content at random times we missed out on a great many opportunities when people may have been receptive to it. The differences in these patterns showed up mainly in their specific activities and very little in the other attributes of context that can be detected by sensors. Further, it appears that people are particularly willing to request content when they are waiting, and especially when they are in microwaiting moments within some other activity.

Preferences for Push and Pull

During the exit interviews, we asked participants if they would prefer to have information automatically delivered to their mobile device (Push) or to trigger it themselves (Pull) or some combination (Depends). Their responses are shown in Table 3.

While the numbers are too small for a reliable statistical analysis, they suggest that requesting content was seen as more attractive than receiving it automatically, although to some extent people preferred what they had gotten used to. In the interviews, people raised concerns about irrelevant content and wanted to avoid *"that feeling of junk mail."* One person said, *"I much prefer to be in control, just because of the technologies presently being imperfect in predicting even remotely what kind of content I wish."*

Condition	Push Only	Pull Only	Depends	Total
Pull	1	10	5	16
Push	6	4	6	16
Total	**7**	**14**	**11**	**32**

Table 3. Number of participants preferring methods of receiving content on their phones

This issue of control came up repeatedly in the interviews. Some were concerned about content arriving at inappropriate times. *"I'd rather trigger it myself. I'd like to have it available to me, because I'm a control freak.… When I'm walking my dogs, I don't want something to start playing and distract me if I'm crossing the street, for safety reasons."* Another cited concerns about bothering others. *"If I were with a group of people, I wouldn't necessarily want it chirping. But if I was by myself I'd be pretty amenable because I could always just look at it quickly or ignore it."* One person in the Push condition wanted the ability to control how information was sent. *"I wish there was a button you can put between manual and automatic. So that way if … I'm not going to be doing anything for a whole day, I would put it on automatic. But if I had a schedule to keep, then I would put it on manual."*

DISCUSSION

What do all these findings tell us about the role of context in determining people's receptivity to content while on the go? We would like to highlight four points that should inform future efforts to offer mobile content services, particularly those that aim to push content to people based on their context. We also offer one point regarding the methods used to study people's information needs.

Limitations of Context as a Predictor of Receptivity to Pushed Content

The aspects of people's context explored in this study – location, activity, time of day, use of media, number of companions, noise level, amount of light, and setting – were somewhat helpful in predicting receptivity to information on the go, but they were not sufficient. Our machine learning model, typical of the kind used by context-aware systems, achieved 66% accuracy when predicting receptivity without knowing in advance whether the person would find the content of interest. If the interest rating was known, accuracy improved to 68%.

It is difficult to assess whether this is "accurate enough" for mobile services that notify people of potentially useful content. In a discussion of recommender systems, Herlocker [15] pointed out that user response varies depending on the domain and is affected by other factors, such as how easily one can act on a recommendation. Still, there is reason to believe that people will be wary of an automated system that is not highly accurate. In one study of an in-car system that notified users of upcoming traffic and hazardous conditions, users' trust in the system dropped off rapidly with inaccurate alerts [22]. When the system was accurate 88% of the time, it was perceived as only 77% accurate.

Another study found that people preferred to choose a route manually rather than pressing a button to have one chosen for them, even when they were given feedback that they were not choosing the most efficient routes [10]. The authors manipulated errors and found that people were much more sensitive to the system's errors than their own. This finding is consistent with the basic human bias known as the fundamental attribution error, i.e. that people see the flaws of others as fundamental character attributes while they see similar flaws of their own as due to temporary situational variables [31]. This phenomenon suggests that automated systems need to perform at a very high level for people to trust their accuracy. Indeed, participants in the driving hazards study perceived a 100% accurate system to be only 90% accurate [22].

We found that the most important aspect of people's context in determining their receptivity was their specific location and activity. The importance of location is consistent with other studies' findings, and we believe our finding on activity is a useful contribution. Further, it is critical to be able to make relatively fine-grained distinctions in these attributes. High-level distinctions such as in public or in private locations, or communicating vs. engaged in food-related activities were not sufficient to predict receptivity. The model achieved moderate success only once we were able to distinguish such activities as preparing food vs. eating it, for example, or driving vs. being a passenger in a car. Similarly for location, it was important to be able to distinguish being on a bus from being in a car – lumping them into one "transit" category was not helpful enough. Context-aware systems will need to detect people's mobile context at this level of detail to do better than chance at predicting the right moment to provide content to people on the go.

Microwaiting Moments

Even within specific types of activities, people's receptivity varied more than could be accounted for by our predictive model. People were especially likely to request content while waiting, such as waiting in line or waiting for a bus or train, and so a context-aware system might do well to learn to detect "waiting behaviors." However, even detecting waiting can be tricky, as many of the waiting situations were not these prototypical types of waiting. Instead, they were microwaiting moments, short idle periods that occurred *in the middle of* other activities, such as waiting for the toast to pop up, for a train to pass, or a break in between poker hands. Other times participants initiated waiting moments themselves, as when they took a short break from reading or exercising, or when they chose to distract themselves from an unpleasant situation. Note that we cannot be sure that people will be highly receptive to pushed content during microwaiting moments; it may be that the loss of control over timing reduces receptivity even at convenient moments.

Our data indicate that people's experienced "context" varies moment by moment, even when none of the sensed context factors change. This finding supports the interactional notion of context as something that, as Dourish explains, "cannot be a stable, external description of the setting in which activity arises. Instead, it arises from and is sustained by the activity itself" [11]. Our data support the idea that "contextuality… is an emergent feature of the interaction, determined in the moment and in the doing." As we have seen, the moment is often quite brief and can change much more rapidly and fluidly than other, more objective measures of context.

We should point out that the content used in our study consisted of short pieces of trivia that could be read in a matter of seconds, so it is not surprising that people chose to request it during short breaks. If instead our information had been page-long news articles, participants' information receptivity patterns may have been very different. However, information delivered to mobile devices is often short due to the small screens and the assumption that people on-the-go want concise information. Thus, our use of short snippets is probably representative for many mobile applications, such as recommendations, location-based product alerts, and notifications of nearby friends or changes in traffic patterns [6, 14, 28, 29].

The Need to Provide Control Over Content Retrieval

Our results suggest that the issue of control is as important as offering content that can be absorbed at moments of low engagement. In the interviews, people frequently mentioned wanting control over receiving content. Although people who had been paid to receive content at random times for two weeks were relatively open to getting pushed content, the overall preference was for requesting content. If it was pushed at them, they wanted the option to decide whether and when to accept it.

The fact that people in the Pull condition requested enough snippets to receive the full daily monetary reward only 30% of the time – and in those cases frequently continued to pull beyond that – indicates that the money was not the only factor. Being able to choose when to get snippets, and the fact that they were interesting, also played an important role. This preference for control again echoes the finding of [10], who showed that people preferred to choose routes themselves even when it was much easier to request a route from the system.

Use Context to Improve Content When Requested vs. Predicting the Right Moment

All these points taken together suggest that people may be willing to receive content, even during activities that might at first seem non-interruptible, if it can be done during short microwaiting moments and if the content is engaging. Predicting these moments, however, is a challenging proposition. The goal of mobile services, then, perhaps should not be to try to predict the perfect time to interrupt and the perfect information to deliver in each context. Instead, we might give people a reason to request content at convenient moments and use all the information detected about their context to improve the relevance and quality of the content delivered. For example, someone seeking "dinner ideas" from a grocery store in the afternoon might get several recipe suggestions, whereas the same query from someone standing with friends on a downtown street in the evening might get restaurant suggestions.

We should also recognize that our accuracy in understanding people's context is bound to vary, and that people quickly lose confidence in automated systems after only a few inappropriate alerts [10, 22]. So a principled approach (comparable to [16]) might be that when we have a strong idea of people's context and receptivity, we offload them by responding appropriately to minimal queries or by proactively suggesting useful content. But when we know less about their context, we should be more cautious and rely on people to specify more about their desires to provide the right content in the right way.

Combining Push and Pull Methodologies

Finally, this study offers a contribution to the methodological question of how to study information receptivity. While studies of interruptibility tend to rely on experience sampling, i.e. probing people at random moments and determining their receptivity, information needs studies often use a diary method where they ask people to capture receptive moments and then later explain the context and reasons for the desired content. In this study, we included both "push" and "pull" conditions and found that these two approaches do not tap into the same notion of receptivity. Our machine learning model was able to predict receptivity to *pushed* content at 16% above chance by taking into account the specific type of location and activity, improving to 18% above chance if we knew how interesting the content was to the recipients However that same model performed only slightly above chance (1%) in predicting when people would *request* content.

Comparing the pull and push data from this study helps illustrate the double-sided nature of information receptivity. On one hand, the experience sampling method may greatly underestimate information receptivity, since it can only probe for passive acceptance of content at random times and cannot tap into those fleeting moments when people are willing to get information. On the other, understanding only when people actively seek information cannot help us characterize when they're *not* open to receiving pushed content. Taken together, these two approaches allow us to gain a fuller understanding of the subtle contextual factors that affect people's willingness to receive and request content while on the go.

ACKNOWLEDGMENTS

Paul Stewart, Kurt Partridge, and Alan Walendowski provided invaluable help with the technical systems aspects of this study. We are grateful to the 32 participants who gave us extensive daily reports of their activities.

REFERENCES

1. Bureau of Labor Statistics. American Time Use Study: Activity Lexicon, 2007.
2. Achloptas, D. and Horvitz, E., *Principles of bounded deferral for balancing information awareness with interruption*. 2005, Microsoft Research.
3. Adamczyk, P. and Bailey, B. If not now when? The effects of interruptions at different moments with in task execution. *Proc. CHI 2004*, 1 (2004), 271-278.
4. Bailey, B. and Konstan, J. On the need for attention aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Journal of Computers in Human Behavior*, 22 (2006), 709-732.
5. Brown, B., Sellen, A., and O'Hara, K. A diary study of information capture in working life. *Proc. of CHI 2000*, 2 (2000), 438-445.
6. Campbell, A., et al. The Rise of People-Centric Sensing. *IEEE Internet Computing*, 1 (2008), 12-21.
7. Charmaz, K., *The Grounded Theory Method: An Explication and Interpretation*, in *More Grounded Theory Methodology: A Reader*, B. Glaser, Editor. 1994, Sociology Press: New York. p. 95-115.
8. Church, K. and Smyth, B. Understanding Mobile Information Needs. *Proceedings of Mobile HCI*, 1 (2008), 493-494.
9. Czerwinki, M., Horvitz, E., and Wilhite, S. A diary study of task switching and interruptions. *Proc. of CHI 2004*, 1 (2004), 175-182.
10. De Vries, P., Midden, C., and Bouwhuis, D. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, 58 (2003), 719-735.
11. Dourish, P. What We Talk About When We Talk About Context. *Personal and Ubiquitous Computing*, 8 (2004), 19-30.
12. Drossos, D. and Giaglis, G.M. Mobile Advertising Effectiveness: An Exploratory Study. *IEEE International Conference on Mobile Business* (2006)
13. Fogarty, J., et al. Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction*, 12 (2005), 119-146.
14. Gaonkar, S., et al. Micro-Blog: Sharing and Querying Content Through Mobile Phones and Social Participation. *Proc. of MobiSys*, 1 (2008), 174-186.
15. Herlocker, J., et al. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22 (2004), 5-53.
16. Horvitz, E. Principles of mixed-initiative user interfaces. *Proc. of CHI 1999*, 1 (1999), 159-166.
17. Horvitz, E. and Apacible, J. Learning and reasoning about interruption. *Proc. of ICMI '03*, 1 (2003), 20-27.
18. Huynh, T., Blanke, U., and Schiele, B. Scalable recognition of daily activities with wearable sensors. *Proc. 3rd International Workshop on Location- and Context-Awareness* (2007)
19. Iqbal, S. and Bailey, B. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *Proc. of CHI 2005*, 1 (2005), 1489-1492.
20. Iqbal, S. and Bailey, B. Effects of intelligent notification management on users and their tasks. *Proc. of CHI 2008*, 1 (2008), 93-102.
21. Iqbal, S. and Horvitz, E. Disruption and recovery of computing tasks: Field study, analysis, and directions. *Proc. of CHI 2007*, 1 (2007), 677-686.
22. Jonsson, I., Harris, H., and Nass, C. How Accurate Must an In-Car Information System Be? Consequences of Accurate and Inaccurate Information in Cars. *Proc. of CHI 2008*, 1 (2008), 1665-1674.
23. Kapoor, A. and Horvitz, E. Experience sampling for building predictive user models: A comparative study. *Proc. of CHI 2008*, 1 (2008), 657-666.
24. Logan, B., et al. A Long-Term Evaluation of Sensing Modalities for Activity Recognition. *Proceedings of Ubicomp 2007* (2007)
25. Naaman, M., Nair, R., and Kaplun, V. Photos on the Go: A Mobile Application Case Study. *Proceedings of CHI 2008*, 1 (2008), 1738-1748.
26. Nagel, K., Hudson, J., and Abowd, G. Predictors of availability in home life context-mediated communication. *CSCW '04*, 1 (2004), 497-506.
27. Ogata, H. and Yano, Y. Context-Aware Support for Computer-Supported Ubiquitous Learning. *Proc. of Wireless and Mobile Technologies in Education*, 1 (2004), 27-34.
28. Raento, M., et al. ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. *Pervasive Computing*, 1 (2005), 51-59.
29. Rao, B. and Minakakis, L. Evolution of Mobile Location-based Services. *Communications of the ACM*, 1 (2003), 61-65.
30. Ratsimor, O., et al. eNcentive: A framework for intelligent marketing in mobile peer-to-peer environments. *Proceedings of ICEC 2003*, 1 (2003), 87-94.
31. Ross, L., *The intuitive psychologist and his shortcomings: Distortions in the attribution process*, in *Advances in experimental social psychology*, L. Berkowitz, Editor. 1977, Academic Press: New York.
32. Senseable City. http://senseable.mit.edu.
33. Shen, X. and Chen, H. An Empirical Study of What Drives Consumers to Use Mobile Advertising in China. *Proc. of The 3rd International Conference on Grid and Pervasive Computing*, 1 (2008), 158-163.
34. Sohn, T., et al. A Diary Study of Mobile Information Needs. *Proc. of CHI 2008*, 1 (2008), 433-442.
35. Takamae, Y., et al. Estimating human interruptibility in the home for remote communication. *Proc. of CHI 2006*, 1 (2006), 1397-1402.
36. Takemae, Y., et al. Analysis of human interruptibility in the home environment. *Proc. of CHI 2007*, 1 (2007), 2681-2686.
37. Tsang, M., Ho, S.-C., and Liang, T.-P. Consumer Attitudes Toward Mobile Advertising: An Empirical Study. *International Journal of Electronic Commerce*, 8 (2004), 65-78.
38. Vastenburg, M., Keyson, D., and de Ridder, H. Interrupting people at home. *IEEE International Conference on Systems, Man and Cybernetics*, 1 (2004), 59-64.