Photographing Information Needs: The Role of Photos in Experience Sampling Method-Style Research

Zhen Yue
University of Pittsburgh
135 North Bellefield Ave.
Pittsburgh, PA 15213, USA
zhy18@pitt.edu

Eden Litt
Northwestern University
2240 Campus Drive Evanston
Chicago, IL 60208, USA
eden.litt@u.northwestern.edu

Carrie J. Cai
MIT CSAIL
32 Vassar Street
Cambridge, MA 02139, USA
cjcai@mit.edu

Jeff Stern
Elon University
100 Campus Drive
Elon, NC 27244, USA
jstern@elon.edu

Kathy Baxter, Zhiwei Guan, Nikhil Sharma, George Zhang
Google Inc.
1600 Amphitheatre Pkwy, Mountain view, CA 94043, USA
{kathyb*, zguan, nikhilsh, georgez}@google.com
*corresponding author

ABSTRACT
The Experience Sampling Method (ESM) enables researchers to capture information about participants' experiences in the moment. Adding an end-of-day retrospective survey also allows participants to elaborate on those experiences. Although the use of photos in retrospective interviews and surveys for memory elicitation is well known, little research has investigated the use of photos in ESM studies. As smartphone adoption increases facilitating ESM studies and making photo sharing easier, researchers need to continuously evaluate the method and investigate the role of photos in such studies. We conducted a large-scale ESM and retrospective survey study via Android smartphones with more than 1,000 US participants, and analyzed participants' photo submissions, including how photo use correlated with participants' data quality and what, if any, value photos added for researchers. Our study sheds light on the role of photos in ESM and retrospective studies that researchers can reference when constructing future study designs.

AUTHOR KEYWORDS
Experience sampling method; photo-elicitation; information need; retrospective study method.

ACM CLASSIFICATION KEYWORDS
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces (Evaluation/Methodology).

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2473-1/14/04.
http://dx.doi.org/10.1145/2556288.2557192

INTRODUCTION
The Experience Sampling Method (ESM) [1] refers to a method for collecting data from a participant in the natural context of everyday life. In an ESM study, participants are reminded randomly during fixed windows of time and asked what they are doing in that moment. It is based on the work of Edmund Husserl's "pure phenomenology," which says that the only things we can really know are the events represented in our individual streams of consciousness [13]. The ESM was also influenced by William James who stated that a person's life can be seen as the sum of all of his or her experiences accumulated over a lifetime [16]. The ESM is designed to be a reliable measure of events over time. Compared to a survey, diary and other self-reported study methods, the ESM is less susceptible to subjective recall errors because the focus is on the participant's immediate experience [14]; however, it can be disruptive to participants' current activities. One way of reducing disruption is to ask participants to enter briefly what they are doing when alerted and then later on have them fill out a more extensive survey. To aid in participant recall, participants are sometimes encouraged to take photos or videos for later review in retrospective interviews or surveys [2]. Russell and Oren [27] found in a study on search behavior that cuing participants with their screen captures aided in their recall accuracy. With smartphone use increasing in popularity, researchers are also turning to the technology for ESM studies making photo submissions for retrospective purposes a viable study design option. Despite the proven value of photos in other types of studies, there has been minimal work evaluating the role of photos in ESM studies. In particular, most studies focus on how photos help participants recall events, while the value of photos to researchers is neglected.

[1] The method is sometimes referred to as Ecological Momentary Assessment (EMA) when used in the medicine domain [29].

In this paper, we discuss a large scale ESM study with retrospective surveys conducted to explore people's daily information needs with the goal of identifying innovation opportunities for a search engine. Previous studies using search logs analysis [17] provide us with a landscape for what people use a search engine; however, people do not solve all information needs online. Logged search queries may be just a small fraction of people's daily information needs. Diary studies [30] are often the alternative to log analysis for understanding information needs.

We used the ESM combined with a retrospective survey because we wanted a more reliable way of capturing people's information needs in a natural context. In this study, participants provided text descriptions of their information needs throughout the day for five days and were encouraged to submit photos if it would help better describe their needs. We recognize that taking a photo for every information need could be burdensome or socially inappropriate at times, so we made it an optional activity [4]. In addition, participants were asked to complete a retrospective survey at the end of the day to describe more about their information needs. While the original purpose of this study was to collect peoples' information needs, the focus of this paper is on the evaluation of the methodology and the role that photos played in the study design more generally. Who submits photos? Do participants stay on task and submit relevant photos for the primary goals of the study? Do photos help participants provide higher quality responses during the retrospective parts of the study [27], or do they interfere with the participants' goals and the study's goals? Do researchers understand the photo submissions and find them useful or might they only be useful to participants? In this paper we address the aforementioned challenges through the following research questions:

RQ1. Who submits photos and when?

RQ2. Do photos help participants provide higher quality data without interfering with the participants'/study's primary goals (e.g., their information seeking)?

RQ3. Do photos help researchers understand participants' responses (e.g., their information needs) better?

RELATED WORK

The ESM has grown in popularity since Csikszentmihalyi, Larson and Prescott published a report on one of the first and most well-known ESM studies in 1977 [5]. The method is revered for its ecological validity and reduction in memory bias as well as its ability to capture contingent observations and within-person processes [28]. The ESM is commonly used in psychology to study concepts surrounding experiences of the self like mind-wandering [31], work stress and satisfaction, and relationship satisfaction. Researchers have also used the ESM to study experiences with games [7], ubiquitous computing systems [4], and programming software [15]. All of these studies are able to collect data from people in their natural environment over an extended period of time. Most ESM studies last one to two weeks [28] during which participants are 'pinged' (alerted) 2-12 times throughout the day at random times and asked to report their behavior or mood. Early studies were limited by technology, requiring participants to either set an alarm or receive a phone call at home. In these studies, participants would either anticipate the ping, or researchers were limited to only studying participants in their homes. Advances in technology have eliminated these early methodological problems and, now, researchers are using smartphone technology to help facilitate ESM studies. Smartphone applications such as Maestro [22] and the Personal Analytics Companion (PACO) [2] alert the participant, present them with a set of questions, and automatically log the data. These applications eliminate the need for participants' access to pen and paper or additional devices, potentially making the methodology less intrusive to participants' daily routines.

However, the ESM has some drawbacks. The act of being pinged frequently throughout the day may be intrusive and the people that agree to participate in such studies may lead to a self-selection bias [28]. Some studies have suffered from low completion rates; studies that sample 8 or more times per day over one to two weeks get a 50-80% completion rate [7]. These drawbacks aside, research has shown that the ESM is a viable method for collecting behavioral or emotional activity directly from people in their natural context over time [28]. However, questions still exist about what collected information is most useful for researchers and participants. In particular, the value of photo submissions in ESM studies, including how to collect and utilize photos is unclear.

Photos may be particularly relevant to ESM studies because of *elicitation*, or their ability to aid participants in providing retrospective interview and survey responses [12]. For example, Collier [3] observed that when a researcher used photos during an interview, participants' responses tended to be longer and more pointed than those in the control group. Others have used photo elicitation to enhance memory and learning. For example, Lee and Dey [19] designed a life-logging system with automatic photo and audio-capture to assist people with memory impairments by cuing them to remember details from their daily experiences [19]. Photo elicitation may also be a positive anchoring tool in educational contexts [6]. Not only can photos stimulate discussion and learning via vivid and lucid imagery [9], but they may also serve as effective verbal prompts, potentially increasing recall rates [6].

Photos have been used in some diary studies [2,8] and ESM studies [4,14] to help participants capture their experiences and recall memories. For example in a study evaluating ubicomp applications, researchers found photo submissions

[2] https://quantifiedself.appspot.com/main.jsp

helpful in highlighting what was important to participants however, the researchers did not further analyze the photo's role or usefulness. Similarly, Intille and colleagues [14] prototyped a non-intrusive image-based ESM that automatically took photos of participants' contexts, but no formal study was conducted to evaluate the method. Gabridge and colleagues [8] evaluated users' information-seeking behaviors in a photo diary study on library systems, but focused on the information needs found rather than the role of photos in eliciting those needs. Despite the usage of photos in research studies, little research has systematically investigated the extent to which photo taking is related to eliciting responses from participants or helping researchers understand participant responses better.

As smartphone use popularizes, and photo sharing online also continues to gain popularity [25], increasing the likelihood of photo-sharing uptake in studies, more researchers are considering incorporating photos in their ESM-style research designs. However, more inquiry is needed to see if there may be biases in who actually participates and submits photos. For example, background factors like age and gender tend to relate to who shares photos online more generally; women are more likely to engage with photo-sharing services online than men, and tend to upload more images [e.g., 21,32]. Young adults are also more likely to post their own photos online than older adults [26]. Additional factors related to one's technological experience are also associated with who shares online [10], including if and how people share their photos [23]. In this paper, we examine whether photo submissions in ESM studies are randomly distributed throughout the sample, or whether there are also similar systematic patterns among photo-sharers online and photo-sharers in research studies online, and the implications behind such potential patterns.

The lack of a systematic evaluation of participants' photo use and focus on the value they have for researchers, combined with the relatively small number (<50) of participants in prior photo-related studies, leaves many questions unanswered. Some researchers have debated the potential shortcomings of the use of auto-photography and elicitation in research [33]. Allowing participants to determine what to photograph also limits the researcher's control over what information can be elicited [24]. A study comparing three media (photos, audio, and tangible artifacts) in diary studies [2] found that photos lead to more specific recall than the other two, but only 11 participants were included. Our study incorporates data from a large-scale ESM study of more than 1000 people, systematically investigates photo submission as it relates to the quality of responses, and probes the extent to which such responses are in practice useful to researchers.

Based on prior research, it is clear that the use of photos in ESM and retrospective studies is an important topic, but many questions remain. The goal of this article is to address this gap in the literature.

METHODOLOGY

The ESM and end-of-day study design

We recruited more than 1,000 Android phone users across the US through a vendor as well as through our own participant database. The study ran between March and May 2013, and was conducted in five waves that each lasted five days, with 200-250 participants per wave. Fifty-two percent of the participants were male and participants ranged in age from 18 to over 60. Participants represented 46 out of 50 states in the country as well as Washington, D.C.

Participants were asked to install the Android app "PACO" from the Google Play Store [3] on their smartphone. Participants that successfully installed the app were sent notifications randomly 8 times a day (between 9am and 7pm in the participants' local time) and asked to complete a form about their information needs. The form asked basics about the information need including what it was ("What"), how important it was ("Importance") and how urgent it was ("Urgency"). For the "What" survey item, we instructed participants to describe their most recent information need using a sentence and provided an open text field. The "Importance" question was a single-selection question with a 5-point Likert scale and the "Urgency" question was a single-selection question with a 7-point Likert scale.

During each notification (beneath the "What" question), participants also had the opportunity to submit a photo with their entry. This was optional. The instructions stated, "When to include a photo? - Whenever it gives us insight about the information you needed and why." Participants had one hour to submit information before the notification timed out and was marked as "missed." Participants also had the option of manually submitting information without having received a notification, whenever they had a need they wanted to share.

At the end of the day, participants were sent a final notification to complete a survey on their desktop or laptop. The retrospective survey showed participants their text and photo submissions (if applicable) from that day and asked more information about their needs including why they needed the information ("Why"), how much of the information they were able to find that day ("Success") and how easy or difficult it was to find the information ("Ease"). The participants were required to answer all of the questions for each of their submitted information needs. For the "Why" survey item, we instructed participants to describe why they wanted to know the information using a whole sentence and left an open text field. The "Success" question was a single-selection question with a 5-point Likert scale, and the "Ease" item was a single-selection question with a 7-point Likert scale. The survey asked additional questions related to information seeking such as

[3] https://play.google.com/store/apps/details?id=com.pacoapp.paco

sources participants used to look for the information, but we do not analyze these in the current paper.

Participants were asked to respond to at least 5 of the notifications per day and complete each end-of-day survey for 5 days. The amount of compensation was based on industry standards for a 5-day rigorous engagement. If they completed 3 responses and the end-of-day survey for 3 days, they received $150 in incentives. If they responded to 5 or more notifications plus the end-of-day survey for 5 days, they received $200 in incentives.

Quantitative analysis method

The analysis unit in this study is a response from a participant that describes one information need. In the following part of this paper, we refer to this as a "DIN" (Daily Information Need). Excluding notifications that did not get responses, we have 33,180 DINs in the original dataset. If a participant responded to a notification but did not complete the end-of-day survey, the DIN was marked as incomplete. If the participant responded "nothing," "no need," etc., or if the response was about the study itself, the DIN was marked invalid. After removing the incomplete and invalid responses, there were 25,368 DINs from 1,013 participants.

Because photo submissions were optional, not all DINs were associated with a photo. Among all the DINs in the cleaned dataset, 889 (3.5%) DINs were associated with a photo. Therefore, two types of participants were identified for our analysis: 1) Photo-sharers, or participants who submitted at least one DIN with a photo; 2) Non-photo-sharers, or participants who did not submit any DINs with a photo. However, photo-sharers did not always submit photos with their DINs, so we categorized the whole dataset into three groups (shown in Table 1). The first group (G1) includes DINs from non-photo-sharers. The second group (G2) includes DINs from photo-sharers that did not have a photo submission associated with it. The third group (G3) includes DINs from photo-sharers that have associated photos. Comparing G1 and (G2+G3) allows us to analyze any differences between photo-sharers and non-photo-sharers, whereas comparing G2 and G3 enables us to examine any differences between DINs without photos and DINs with photos.

	From non-photo sharers (G1)	From photo sharers but without photos (G2)	From photo sharers and with photos (G3)
# DINs	17,182	7,297	889

Table 1: Three groups of DINs.

In order to answer RQ2 (Do photos help participants provide higher quality data without interfering with the participants'/study's primary goals (e.g., their information seeking)?), we measured both data quality and the "Ease" and "Success" of participants' original primary goal, information seeking. One set of measurements used for evaluating the data quality is the rate of incomplete and invalid DINs. From the original dataset of 33,180 DINs, we removed about 7,000 incomplete and 1,000 invalid DINs. We compared the incomplete and invalid rates among the three groups with the assumption that higher incomplete and invalid rates indicate lower data quality.

To further evaluate data quality, we also measured the length of participant responses to the "What" and "Why" questions. We chose to focus on response length because previous research has shown word count to be an effective quality measure [1], typically allowing for more opportunity to understand participant responses. We thus used the number of words as one measurement. Because many responses were submitted through mobile interfaces, we note that each additional character adds extra effort for the participants. Therefore, we also used the number of characters as another measurement of data quality. In the analysis, we assumed that the higher the word or character count, the higher the quality.

Participants' responses to the "Success" and "Ease" questions were used to measure the success and ease of their information seeking for each DIN. For the analysis, we assume that the higher the ratings, the more successful participants were at solving their information need and the easier it was for them to find the information.

Qualitative analysis method

To determine the extent to which photos can provide additional information to researchers, we recruited 12 researchers to manually code all 889 photos in the dataset. The photographs were divided into four subsets of 222 or 223 photos, which were randomly assigned to the researchers such that each photo was ultimately coded by three researchers. The coding process was conducted using a Web interface. To help researchers isolate the content they could surmise from text alone, the webpage first presented the "What" and "Why" responses without the photo. Once the coder clicked "Show image," the photograph appeared along with two questions for the coder. The first question captured whether the photo was relevant or not and asked "Is this photo RELEVANT to the information need that he/she wrote above?" The second question measured the photo's usefulness to researchers and asked, "Does this photo help you understand more about the information need BEYOND what he/she wrote above?" Each question had only "Yes" or "No" as answer options, but researchers were allowed to skip a maximum of 5% of trials in rare cases when they were unable to make a decision. If researchers mistakenly answered "No" to relevance but "Yes" to usefulness, they were given a pop-up warning that prevented them from advancing to the next trial until they reevaluated their choices.

Prior to coding, researchers were trained on standards of relevance and usefulness, and arrived at a consensus regarding the usefulness of a photograph. A photo was deemed useful if the researchers thought the photo helped

them understand the participants' information beyond what was written in the text alone. To determine overall relevance and usefulness, answers for each photo were aggregated via a majority vote among the three coders. If a photo was deemed relevant in the first question by at least two raters, we then determined its usefulness using a majority vote on the second question. We treated a skipped photo as a vote for "Skip," so if more than one rater skipped the photo, we removed it from the qualitative analysis.

RESULTS

In this section, we report results for each of the three research questions.

Who submitted photos and when?

Photo-sharer versus Non-photo-sharer

While participants were not required to submit photos during the study, almost a third (30.80%), or 312 participants submitted at least one photo. Participants submitted 889 photos during the study, accounting for roughly 3.50% of DINs. On average, photo-sharers uploaded approximately three photos during the study (M=2.84). Table 2 showcases information on photo-sharers in comparison to non-photo-sharers.

		Photo-sharer	Non-photo-sharer	Statistic test
Gender	**Female**	170	316	$\chi2$=7.28 **p =0.007**
	Male	142	385	
Age	**18-23**	39	128	$\chi2$=11.65 **p =0.02**
	24-30	83	222	
	31-40	109	214	
	41+	80	135	
Mobile phone usage	**Low**	80	199	$\chi2$=0.91 p =0.634
	Medium	146	310	
	High	86	192	
Phone Search frequency	**Low**	106	274	$\chi2$=4.55 p =0.103
	High	205	427	

Table 2: Photo-sharer vs non-photo-sharer (some participants did not provide responses to some of these questions).

Upon examining photo-sharers in comparison to non-photo-sharers, we find photo submissions are not randomly distributed among participants, but rather certain people are more likely to submit photos than others. Similar to photo sharing online, females were more likely to submit photos in this ESM study than males. More than a third of females (34.97%) submitted photos in comparison to just over a quarter of males (26.94%). However, in contrast to photo sharing online, younger adults (18-23) were less likely to submit photos with their DINs than those over 40 years of age. Less than a quarter (23.35%) of the participants aged 18-23 submitted photos compared to over a third (37.2%) of participants aged 41 and older. This result holds true even when controlling for gender effects.

In addition, we do not see any correlation between who submitted photos and self-reported technology use, such as how often participants use their mobile phone and how often they generally search for information. For example, we see no significant difference of photo sharing between people who use their mobile phones frequently in comparison to those who use their phones less often.

When photos were submitted

To get a better sense of the photo submission timeline, we analyzed when participants submitted photos. In general, participants seemed to be more participatory in the beginning of the study as they were more likely to respond to notifications sent to them and more likely to include a photo on the first day of the study in comparison to the other days. However the decay rate was starker for the percent of DINs with photos. When we compare photo-containing DINs with the total number of DINs submitted per day, we still observe that participants were more likely to submit photos on the first day. A total of 272 photos were submitted on the first day, accounting for 5% of the DINs submitted on the first day, compared with only 2.9% on the final day of participation (χ^2 = 50.219, df = 4, p < 0.01). Post-hoc analyses suggest that the participants submitted a higher rate of DINs with photos on the first day. While participants still submitted photos on all five days, there may have been a novelty effect with photo submissions that diminished after the first day.

	Day1	Day2	Day3	Day4	Day5
DINs w/ photos	272	178	146	144	145
Total DINs	5,410	5,130	4,959	4,820	5,027

Table 3: DINs for each day.

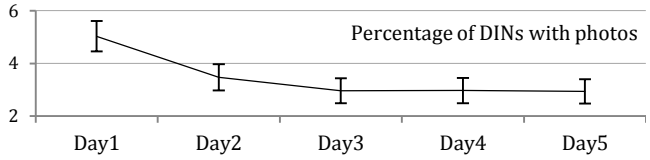

Figure 1: Percentage of DINs with photos for each day.

We did not find any significant differences using chi-square tests in reported "Urgency" or "Importance" of the information needs between DINs with photos and DINs without photos. Hence, our findings did not suggest a correlation between photo submission and urgency or importance of the information needs.

DIN quality and photo interference

A goal of any study is to obtain high quality data without placing undue burden on participants. The purpose of RQ2 was to discover whether photo submissions were associated with higher quality data, (e.g., DINs) and, if so, whether submitting a photo could interfere with the participant's primary goals, in this case, the information seeking process.

DIN quality

Based on prior research on photo elicitation, we hypothesized that DINs with photos were less likely to be

incomplete DINs because photos might trigger the participant's memory while taking the end-of-day survey. Our results (as shown in Table 4) indeed show that DINs with photos have the lowest rate of missing an end-of-day survey (17.1%) and the lowest rate of invalid DINs (1.8%). Compared to the DINs from non-photo-sharers, DINs from photo-sharers but without photos actually had the highest incompletion rate (22%). Chi-square analyses highlighted that the overall differences on both incomplete and invalid rates were significant among the three groups. Post-hoc analyses further indicated the differences on incomplete rates between any two groups were significant, and the invalid rate for DINs with photos was significantly lower than the other two groups. Overall, these results indicated DINs with photos were more likely to be higher quality in comparison to DINs without photos, while DINs from photo-sharers without photos were less likely to be higher quality than DINs from non-photo-sharers.

	From non-photo-sharers	From photo-sharers (no photos)	From photo-sharers (with photos)	Statistic test
Incomplete DINs (%)	20.5	22	17.1	$\chi 2$=19.58 p<.001
Invalid DINs (%)	2.9	2.5	1.8	$\chi 2$=19.04 p<.001

Table 4: Percentage of incomplete and invalid DINs.

Furthermore, we found significant differences among the three groups in terms of the length of "What" and "Why" responses. We predicted that DINs with photos would have shorter "What" responses because participants may have used photos as partial replacement for text, and we hypothesized that "Why" responses in the EOD survey would be longer because photos would help them recall and describe their information needs more comprehensively.

	From non-photo-sharers	From photo-sharers (no photos)	From photo-sharers (with photos)	Statistic test
# word in "What"	8.00	8.23	9.13	F=32.53 p<.001
# char in "What"	42.17	43.10	47.16	F=24.48 p<.001
# word in "Why"	13.54	14.88	16.27	F=90.55 p<.001
# char in "Why"	67.89	74.77	81.66	F=87.79 p<.001

Table 5: Length of "What" and "Why".

Our results (as shown in Table 5) show that DINs with photos had both the longest "What" (9.13 words or 47.16 characters) and "Why" (16.27 words or 81.66 characters) among the three groups. DINs from non-photo-sharers had the shortest length of "What" (8 words and 42.17 characters) and also the shortest length of "Why" (13.54 words and 67.89 characters). One-way ANOVA tests show that the overall differences among the three groups were significant for each of the four measurements on length. Post-hoc analyses with Bonferroni correction also indicate the differences between any two groups were significant for each of the four measurements on length. Hence, DINs from non-photo-sharers tended to have shorter responses than DINs from photo-sharers. More importantly, among all the DINs submitted by photo-sharers, those with photos were associated with longer responses than those without photos.

The above analysis shows that submitting a photo correlated with higher data quality, operationalized in terms of response length and valid DINs.

Photo interference

Because photo submission required extra effort from the participant, an important question to address is whether submitting photos could affect the participant's original primary goal at hand, finding information. We analyzed two self-reported questions about participants' "Ease" and "Success" regarding finding information for DINs from photo-sharers, including those without photos (G2) and those with photos (G3). We excluded DINs from non-photo-sharers (G1) to control for the possibility that "Ease" and "Success" of finding information could be highly related to the search expertise of participants. Because G2 and G3 were DINs from the same set of photo-sharing participants, we can mitigate the possible effect of search expertise. Because the responses for these two questions were ordinal and the distributions were skewed, we use Wilcoxon Signed Rank Test. The results (as shown in Table 6) indicate that there was no significant difference on either "Ease" or "Success" between G2 and G3. Hence, we find no evidence that submitting photos had a negative (or positive) impact on the participants' ability to fulfill their primary goal.

	From photo-sharers (no photos)	From photo-sharers (with photos)	Statistic test
Median "Ease"	3	3	W=2445314 p = 0.49
Median "Success"	2	2	W=2460039 p = 0.10

Table 6: "Ease" and "Success" of finding information

Relevance and usefulness of photos

Our third research question investigated whether the photos are helpful to researchers. Researchers coded 882 photos (skipping seven photos) into one of three categories: irrelevant, relevant but not useful, and relevant and useful (For confidentiality and privacy, we provide simulated but representative photo examples for each category as shown in Figures 1, 2 and 3). The Fleiss' Kappa inter-rater reliability (IRR) for the relevance question was 0.53 and the IRR for the usefulness question was 0.35, which indicated

that researchers had moderate agreement on the relevance and usefulness of photos. The final category of each photo was determined by majority voting. While these are acceptable IRRs [18], we see that even with training, researchers still had some difficulty in agreeing upon the usefulness of the photos for data analysis.

Irrelevant photos

Researchers rated 8.9% of the photos as irrelevant, indicating the researchers did not think that particular photo aligned with the participant's DIN text. A closer look at the photos revealed that many were unrelated images of the individuals, images that contained shots of their immediate context, and photos that were undecipherable. For example, one need was "[Wanted] to know if there was an evening Zumba class at any Mountainside Fitness location" and was accompanied by a picture of the participant holding a pen. While these irrelevant photos may have been useful to the participants in triggering their memories later on, the researchers found them irrelevant, and sometimes even distracting, when understanding the information need.

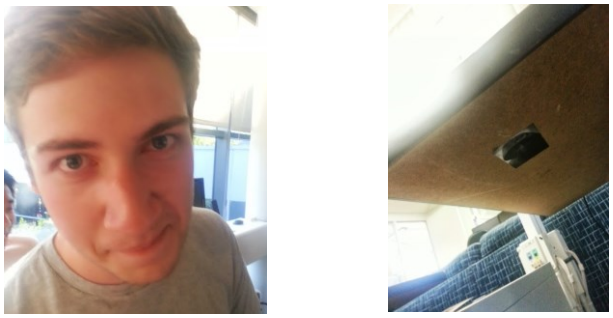

Figure 2: Left: An example of an irrelevant photo of the participant (What: How many days until my trip? / Why: I'm going on a vacation soon and was thinking about when it started); Right: An undecipherable photo coded as irrelevant.

Relevant but not useful photos

Researchers marked the majority of photos as relevant, but not providing any new information (*56% of total*), indicating that while oftentimes the photo was related to the need, it did not provide a deeper understanding. One common theme among this group of photos was a capturing of the information-seeking *tool* itself. For example, a DIN was, "why do my hands burn after putting on lotion?" The accompanying photo was of a search engine results page with the stated need as a search term. It merely signaled how the participant sought to fulfill the need, but it did not clarify or elaborate on the need itself. This was not useful because participants already reported the source they used for each DIN in the end-of-day survey. Furthermore, some photos contained a relevant context, but they were often too general to be useful. For example, one information need stated, "Is the horse hostile to humans?" and the photo was of a horse very far away making it impossible to recognize the horse's breed. A frequent need participants had was about weather conditions ("Is it going to be sunny tomorrow?") and the accompanying photos were often of the current weather itself. Other photos were too blurry or low quality to be useful for the researcher.

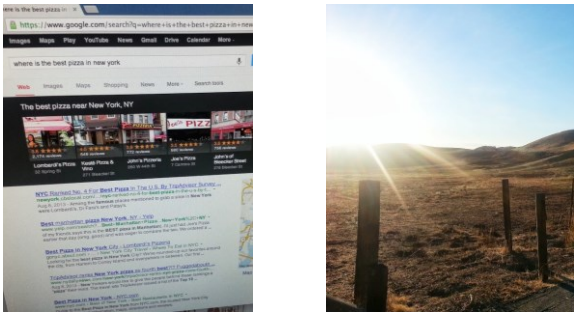

Figure 3: Left: An example of a relevant but not useful photo of an information-seeking medium (What: Where is the best pizza in New York? / Why: I wanted to get pizza for lunch and was trying to decide where to go); Right: A relevant but not useful photo about weather conditions (What: Is it going to be sunny tomorrow? / Why: I wanted to plan for a picnic tomorrow).

Relevant and useful photos

The final category, which comprised 35.1% of the dataset, are photos that were relevant to the DIN text *and* provided the researcher with additional information about the participant's need. This type of photo was most helpful to the researcher. Common themes emerging from this group were photos that clarified non-specific nouns (e.g., text said "this fish"; photo showed the type of fish in question), disambiguating nouns (e.g., text said "keyboard"; photo showed a computer keyboard rather than a musical keyboard). While it is possible that participants would not have used ambiguous or non-specific nouns had they not included the photo, results from the quantitative analysis indicate that the "What" and "Why" text is longer (and possibly more specific) among DINs with photographs.

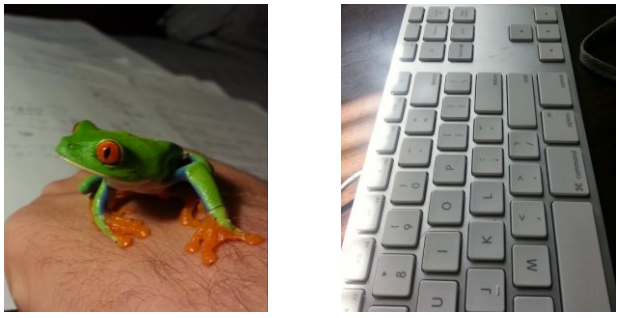

Figure 4: Left: An example of a relevant photo that disambiguates the type of 'pet' referenced (What: What does this animal eat? / Why: My friend showed me his pet and I was curious what it ate.); Right: A relevant photo that provides more information about the style of keyboard and disambiguates it from a musical keyboard (What: Where can I buy a new keyboard? / Why: My old keyboard broke).

Lastly, some photos in this category provided new information to the text by giving more nuance to the information need. For example, a need stated, "My granddaughters are staying with me this summer. I am

looking for crafts to keep them busy." The photograph showed two young girls, identifying the approximate age and number of granddaughters.

Overall we found that while participants submitted relevant photos in the study, for the purposes of our study, the majority of photos did not help researchers understand the participants' needs beyond what they had already included in text responses.

DISCUSSION

Although the ESM has been used as a research method for decades, as smartphone adoption rises and access to potential participants becomes easier through apps like PACO and Maestro, ESM-related studies are also becoming more prevalent. As researchers increasingly use this method for data collection, it is important to evaluate its effectiveness on various levels. Using results from an ESM study of more than 1,000 participants, we investigated whether any systematic biases exist between photo-sharers and non-photo-sharers, how photos relate to participants' data quality, and whether or not photo submissions are helpful for researchers.

In our study design in which photo submissions were optional, we found that photo submission was not common. Even within the third of participants who shared at least one photo during the study, photo submissions only accounted for a small percentage of their responses. Furthermore, we found that some people were more likely to submit photos than others. Females were more likely to share photos with the researchers than males. This finding echoes similar past research, which suggests that because of differing communication patterns and desires, women may be more likely to participate online and share their photos [34]. Surprisingly, in contrast to photo-sharing trends online, we found that those aged 41 and older in our study were more likely to submit photos than those 18-23 years old, even when controlling for gender. Although research on online photo sharing more generally has found a negative relationship with age [26], such findings may be attributable to people's Internet skills [11], which have been linked negatively with age (see [20] for a review). This trend might not hold among participants in our study, who were Android smartphone users and may thus be more technologically savvy than the average adult. The age patterns may also be linked with a social or psychological variable not explicitly measured in this study, such as time availability, privacy concerns, financial motivations, and conscientiousness. Regardless of the specific explanatory variable, the main conclusion here is that photo-sharers tended to be different from non-photo sharers. This is important for researchers who are designing similar studies involving photo submissions. For example, if researchers only focus on submissions containing photos, they may unintentionally bias the overall findings by systematically leaving out certain people.

Researchers considering incorporating photos into their ESM studies should also keep in mind that photo submissions dropped substantially throughout the study. While this may be a limitation of this particular study, which did not provide explicit feedback to participants regarding how their photo submissions were being used, it is also possible that this novelty/drop-off effect may persist in other multi-day studies, in which participants are eager and compliant in the beginning of the study, but by the end they may drop off due to repetition and fatigue.

Beyond photo submissions being related to participants' background characteristics, we also witnessed greater quality of responses when photos were involved. Participants were more likely to complete the end-of-day survey and more likely to share longer responses about their information needs. Although we did not conduct a systematic text analysis to discover whether longer responses were in fact more articulate, our findings are consistent with prior research suggesting the role of photos in memory elicitation [3,6]. Photos may have helped some participants articulate more details about why they needed the associated information, carrying memory triggers beyond what had been written in the text responses alone.

Although our results provide some evidence that photos may be helpful for participants, we also found that photos may not be as helpful or relevant to researchers and data analysis. In accordance with the instruction on submitting a relevant photo, the overwhelming majority of participants submitted relevant photos. While researchers had more difficulty in objectively determining the usefulness of the photos, they found just over a third of photos helped them understand participants' information needs beyond what they had written. This tended to be particularly true when the photos helped disambiguate participants' nouns, clarified their non-specific nouns, or added more nuance to their information needs. Although researchers found the "selfies" (i.e. photos of the participants) and blurry photos irrelevant, it is possible that these photos may have still helped the participant in responding to the end-of-day survey. In the future, researchers may choose to include the photo-sharing option in their study if they desire longer responses from participants, but whether it is worthwhile for researchers to analyze the entire photo dataset remains an open question. Additionally, researchers may find photos useful for other objectives beyond data analysis such as communicating findings in presentations, or creating personas and use cases for design/product development.

CONCLUSION

Overall, our findings suggest both advantages and drawbacks to photo submissions in ESM- and retrospective survey-related studies. Researchers can take these into account for future work and adjust their study designs based on their own primary goals. For example, on the one hand, requiring or encouraging photo submissions may lead to

more biased samples. On the other hand, photo submissions may be linked with signals of higher quality data like longer and more valid responses. Furthermore, as our study demonstrated, researchers may find at least a subset of photos critical to understanding participants' responses.

While our study raises important insights for researchers incorporating ESM techniques and photos into their studies, it is important to keep in mind this study's limitations. For example, although our research allowed for large-scale data collection in participants' everyday environments, our primary study had a specific set of goals, instructions, compensation, and quality indicators. Future research can explore the applicability of these results and issues in more depth by investigating topics such as how variations in instructions may affect photo submission compliance or how participant interviews may impact photo elicitation.

Researchers and developers creating ESM-style technology can also work to optimize (a) when to encourage photo submissions, and (b) when researchers should access the photo data for analysis. For example, the research technology (e.g., PACO) could alert a user to include a photo after automatically detecting an issue such as after the use of a non-specific noun or ambiguous keyword. During analysis, researchers can in turn use some of the themes identified in this paper to pinpoint text responses that may need to be analyzed in conjunction with photo data. Participants could also aid in the process by manually flagging when their photo is vital to understanding the response. Moreover, since many irrelevant photos captured the individuals' immediate context rather than relating to their information need, researchers may consider providing more feedback on how the photos will be used or allow greater flexibility in when participants can submit photos. These may be particularly necessary if the researcher's priority is for the photos to provide additional information.

Future research should investigate the specific factors that influence whether and when participants choose to include photos from the participants' perspective, such as through in-depth interviews. Furthermore, future work can examine if there are any patterns in the types of responses that are more likely to receive photos. Likewise, a more structural analysis of text responses (e.g., using part of speech tagging) may help shed light on precisely what additional verbal content is being included in longer text submissions. Lastly, while our qualitative study provided insight into researchers' perspective on the usefulness of photos, we recognize that other studies and research opportunities may find the photos more or less useful depending on their specific objectives. While it is clear there are both benefits and drawbacks of incorporating photo submissions in ESM- and retrospective survey-style studies, researchers can utilize these findings when constructing future studies.

ACKNOWLEDGEMENTS

We would like to express our gratefulness to Aaron Sedley, Ann Hsieh, Bob Evans, Kerwell Liao, LaDawn Jentzsch, and Marianne Berkovich for their help in this study. We would also like to thank Dan Russell, Ed Chi, John Boyd and the anonymous reviewers for their valuable feedback and suggestions.

REFERENCES

1. Blumenstock, J.E. Size matters: word count as a measure of quality on wikipedia. *Proceeding of the 17th international conference on World Wide Web - WWW '08*, ACM Press (2008), 1095–1096.
2. Carter, S. and Mankoff, J. When participants do the capturing: The Role of Media in Diary Studies. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, ACM Press (2005), 899–908.
3. Collier, J. Photography in anthropology: a report on two experiments. *American Anthropologist 59*, 5 (1957), 843–859.
4. Consolvo, S. and Walker, M. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing 2*, 2 (2003), 24–31.
5. Csikszentmihalyi, M., Larson, R., and Prescott, S. The ecology of adolescent activity and experience. *Journal of Youth and Adolescence 6*, 3 (1977), 281–294.
6. Fanning, S. Visual Methodologies: Photo-Elicitation in the University Classroom. *The Proceedings of the 10th European Conference on Research Methodology for Business and Management Studies*, (2011).
7. Fischer, J.E. and Benford, S. Inferring player engagement in a pervasive experience. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press (2009), 1903–1906.
8. Gabridge, T., Gaskell, M., and Stout, A. Information seeking through students' eyes : The MIT photo diary study. *College & Research Libraries 69*, 6 (2006), 510–523.
9. Grady, J. Working with visual evidence: an invitation and some practical advice. In *Picturing the social landscape: visual methods and the sociological imagination*. Routledge, London, 2004, 18–32.
10. Hargittai, E. and Walejko, G. The Participation Divide: Content creation and sharing in the digital age. *Information, Communication & Society 11*, 2 (2008), 239–256.
11. Hargittai, E. Beyond logs and surveys: In-depth measures of people's web use skills. *Journal of the American Society for Information Science and Technology 53*, 14 (2002), 1239–1244.

12. Harper, D. Meaning and work: A study in photo elicitation. *Current Sociology 34*, 3 (1986), 24–46.

13. Husserl, E. *Logical Investigations, volume 1*. Routledge, New York, 2001.

14. Intille, S., Kukla, C., and Ma, X. Eliciting user preferences using image-based experience sampling and reflection. *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*, ACM Press (2002), 738–739.

15. Jakobsen, M.R. and Hornbæk, K. Fisheyes in the field: Using Method Triangulation to Study the Adoption and Use of Souce Code Visulization. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press (2009), 1579–1588.

16. James, W. *The Principles of Psychology, 2 vols*. Henry, New York, 1890.

17. Jansen, B.J. and Spink, A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management 42*, 1 (2006), 248–263.

18. Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data. *biometrics*, (1977), 159–174.

19. Lee, M.L. and Dey, A.K. Using lifelogging to support recollection for people with episodic memory impairment and their caregivers. *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, (2008), 1–3.

20. Litt, E. Measuring users' internet skills: A review of past assessments and a look toward the future. *New Media & Society 15*, 4 (2013), 612–630.

21. Madden, M. Online photo services. 2005. http://www.pewinternet.org/Commentary/2005/December/Online-Photo-Services.aspx.

22. Meschtscherjakov, A., Reitberger, W., and Tscheligi, M. MAESTRO: Orchestrating User Behavior Driven and Context Triggered Experience Sampling. *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, ACM Press (2010), 1–4.

23. Miller, A.D. and Edwards, W.K. Give and take: A study of consumer photo-sharing culture and practice. *CHI '07*, ACM Press (2007), 347–356.

24. Noland, C.M. Auto-Photography as Research Practice : Identity and Self-Esteem Research. *Jounal of Research Practice 2*, 1 (2006), 1–19.

25. Pew Research Center's Internet & American Life Project. *SNS and Facebook survey 2010*. Washington D.C., 2010.

26. Rainie, L., Brenner, J., and Purcell, K. *Photos and videos as social currency online*. 2013.

27. Russell, D.M. and Oren, M. Retrospective Cued Recall : A method for accurately recalling previous user behaviors. (2009), 1–9.

28. Scollon, C.N. and Kim-Prieto, C. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies 4*, 1 (2003), 5–34.

29. Shiffman, S., Stone, A.A., and Hufford, M.R. Ecological Momentary Assessment. *Annual Review of Clinical Psychology 4*, 1 (2008), 1–32.

30. Sohn, T., Li, K.A., Griswold, W.G., and Hollan, J.D. 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 (2008), 433–442.

31. Stawarczyk, D., Majerus, S., Michalina Maj, M.V. der L., and D'Argembeau, A. Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, (2011), 370–381.

32. Stefanone, M.A. and Lackaff, D. Reality television as a model for online behavior: Blogging, photo, and video sharing. *Journal of Computer-Mediated Communication 14*, 4 (2009), 964–987.

33. Tagg, J. *The Burden of Representation: Essays on Photographies and Histories*. The University of Massachusetts Press, Boston, MA, 1988.

34. Thelwall, M. Privacy and gender in the social web. In S. Trepte and L. Reinecke, eds., *Privacy Online*. Springer, Berlin, Heidelberg, 2011, 251–265.