Remote Shopping Advice:
Enhancing In-Store Shopping with Social Technologies

Meredith Ringel Morris, Kori Inkpen, and Gina Venolia
Microsoft Research
Redmond, WA, USA
{merrie, kori, ginav}@microsoft.com

ABSTRACT
Consumers shopping in "brick-and-mortar" (non-virtual) stores often use their mobile phones to consult with others about potential purchases. Via a survey (n = 200), we detail current practices in seeking remote shopping advice. We then consider how emerging social platforms, such as social networking sites and crowd labor markets, could offer rich next-generation remote shopping advice experiences. We conducted a field experiment in which shoppers shared photographs of potential purchases via MMS, Facebook, and Mechanical Turk. Paid crowdsourcing, in particular, proved surprisingly useful and influential as a means of augmenting in-store shopping. Based on our findings, we offer design suggestions for next-generation remote shopping advice systems.

Author Keywords
shopping; social information seeking; friendsourcing; crowdsourcing;

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]

INTRODUCTION
Shopping in brick-and-mortar (non-virtual) stores is an everyday occurrence for many people. While sometimes done solo, it is often an activity done in pairs or small groups. We call two or more people shopping together *social shopping*. Motivations for social shopping range from task-related reasons, such as getting assistance with decision-making, to relationship reasons, such as companionship and reinforcing social bonds.

Mobile phones, especially smartphones, make it possible to get some of the benefit of social shopping at a distance. The increasing capabilities of mobile phones and of social

platforms further the possibilities for remote collaboration while shopping, an activity we henceforth refer to as seeking *remote shopping advice*. A recent Pew survey suggests that using mobile phones to seek remote shopping advice is an emerging trend − 38% of U.S. shoppers with cell phones made phone calls seeking shopping advice during the 2011 Christmas holiday shopping season, rising to 46% during the 2012 holiday season [21].

We wanted to explore the potential of using smartphones' expanding capabilities for seeking remote shopping advice, in order to discover opportunities for technological innovation in this space. We used a mixed-methods approach to understand current practices and to explore the potential of emerging social platforms (such as online social networks and paid crowd labor markets) to provide remote shopping advice.

We first report findings from a survey of 200 people, detailing their current and desired remote shopping advice habits. Next, we present the results of a field experiment in which people shopping for clothing shared photos of purchase options to (a) a small set of close contacts via MMS (Multimedia Messaging Service, the multimedia version of SMS), (b) their online social network via Facebook, and (c) a set of paid crowd workers on Amazon's Mechanical Turk [mturk.com]. We report on the performance characteristics of these alternatives, including response speed, volume, quality, utility, and influence, as well as participants' comfort level with each experience. Lastly, we synthesize design guidelines for technologies to support seeking remote shopping advice.

RELATED WORK
Our research was informed by related work on shopping habits and technologies and by work on friend- and crowd-sourced information seeking.

Shopping Habits and Technologies
Shopping is an activity that can have important social components [16, 24], although joint shopping trips can pose challenges (e.g., the case of a teenager embarrassed to be seen shopping with a parent) [16]. This paper contributes to the literature on social shopping experiences by presenting survey data on situations in which solo shoppers use mobile technologies to reach out to others, and experimental data on shoppers' reactions to using various social technologies to receive feedback on purchasing decisions.

Hillman et al. have studied how social technologies impact online commerce, such as when making purchases via mobile phones [9] and when sharing deals from social coupon sites (e.g., groupon.com) [6]. Online product reviews (from friends, or even from strangers exhibiting agreement in large numbers) [9] and product-related discussions on message boards and other shopping-focused digital forums [19] have been shown to influence consumer decisions. Rather than focusing on online commerce, our study complements these by investigating how social technologies can enhance in-store shopping.

Recent survey results indicate that consumers often use mobile phones to take photos of products in stores (often as a memory aid, though sometimes to share with others) [26]; in this work, we consider how the practice of photo-taking while shopping can support novel social interactions.

A few prior systems have prototyped novel social experiences surrounding fashion. Tsujita et al. [27] created the Complete Fashion Coordinator, a system to automatically catalogue the contents of a user's closet, and let them select possible outfit combinations and share their choices with close friends and family for comment. Burton et al. [4] used a variant of the VizWiz [2] system to provide fashion consulting to blind women; the women took photos of their outfits with the VizWiz iPhone app and sent the photos to a dedicated fashion "expert" (a volunteer with a fashion-sense profile similar to the target user), who then sent an opinion about the clothing back to the user. In our research, we examine the potential of several social communications technologies (MMS for receiving feedback from specific close ties, Facebook for receiving feedback from one's broad social network, and crowdsourcing for receiving feedback from strangers) to create a socially-enhanced shopping experience for general users.

There are also several commercial forays into this space, though most seem to have had limited success thus far. For example fashism.com and gotryiton.com are two websites that allow users to post photos of contemplated fashion purchases and receive feedback from other site users. More general purpose apps, like See/Saw [seesaw.com] allow users to friendsource general opinion-seeking questions (about items they have taken photos of), and could potentially be applied to shopping. Some retailers have also begun to integrate social technologies into their stores as a differentiator – for example, Sears stores in Canada have experimented with allowing shoppers to engage in Skype video chat with friends about potential purchases [22]. Our research is one of the first to formally study and report on how social technologies impact in-store shopping experiences, including data on the relative performance of various social resources, how users perceive such experiences, and the extent to which they influence users' preferences. Based on these findings, we suggest guidelines for designing remote shopping advice technologies.

Friendsourced and Crowdsourced Information Seeking

Friendsourced information seeking (asking questions of one's personal contacts) and *crowdsourced information seeking* (asking questions of paid crowd workers) are increasingly common phenomena, due to the availability of enabling platforms (e.g., social networking sites such as Facebook for the former and crowd labor platforms such as Amazon's Mechanical Turk for the latter).

Morris et al. [17] found that, as of 2009, about half of the people they surveyed had engaged in friendsourced information seeking using Facebook or Twitter. They also found that questions seeking opinions and recommendations were predominant. Subsequent work by Lampe et al. [14] revealed that not all people view Facebook as an appropriate venue for information seeking; women, younger people, and those with larger online networks were more likely to engage in this behavior.

Many factors influence the effectiveness of friendsourced information seeking on Facebook, including network size (larger networks reduce answer latency) [25], question phrasing (posts that include question-mark punctuation increase response probability) [25], and having social capital (which increases response probability) [11]. These studies considered text-only inquiries; our study adds to the understanding of friendsourced information seeking by studying the performance of a multi-media information seeking task, in which users posted a photo and accompanying question to Facebook.

Jeong et al. [10] studied the characteristics of naturally occurring (i.e., friendsourced) answers to questions on Twitter and of crowdsourced answers generated for those same questions, and found the overall quality of the two information sources to be similar; Twitter users' reactions to receiving serendipitous crowdsourced replies to their public questions from their system were generally positive [10]. VizWiz Social [3] is an iPhone app that lets blind people take photos of objects in their environment, record audio questions, and post them to a combination of crowd- and friend-sourced answering platforms; however, due to their relatively small online social network sizes and their concerns about appearing dependent to friends, the friendsourcing features of VizWiz Social were rarely used, and crowdsourcing was greatly preferred [3]. Our study adds to this growing body of knowledge about the relative performance of crowd- and friend-sourced information seeking by gauging users' reactions to each for decision making during shopping tasks, and compares the performance of the crowd to that of friends for this task.

Panovich et al. [18] studied the impact of tie strength [6] on Q&A exchanges on Facebook, and found that users found more value in answers from stronger ties; our study allows us to examine the value of subjective information received from "non-ties" (strangers on Mechanical Turk) in contrast to that given by personal connections.

SHOPPING SURVEY

We surveyed U.S.-based teens and adults, aged 15 - 60, on their experience with, and desired use of, mobile phones to connect with others while shopping. Respondents were recruited via Cint Link, a professional recruiting service (they paid participants about $4 per survey completion), and we received 215 responses. Fifteen responses were discarded because of poor-quality answers to the free-text recent-critical-incident description question (e.g., typing random character sequences), which left 200 valid surveys that were analyzed. The survey was conducted over a one week period in March 2013.

Most of our respondents fell into three main age groups: 35-54 (57.5%), 18-24 (15.5%) and 25-34 (15.5%). The breakdown of gender was roughly balanced (48% female and 52% male). Additionally, 60% of those surveyed owned a smartphone (a 2012 Pew survey [20] found that 45% of American adults own smartphones, rising to 68% for those in households earning at least $75,000/yr).

45% of our participants reported that they shop in a traditional "brick-and-mortar" store at least once a week (excepting grocery shopping, which was more common). When asked whether they enjoy shopping in these stores, 58.8% responded positively ("slightly enjoy", "moderately enjoy", "strongly enjoy") on a 7 point scale, 14.5% responded negatively ("slightly dislike", "moderately dislike", "strongly dislike") and 26.6% responded neutral. When asked about shopping with a friend, these percentages were very similar (positive – 59.3%, negative – 17%, neutral – 23.6%).

We asked respondents about the perceived value of receiving feedback from several sources while shopping for clothing. Responses were given on a five-point scale, ranging from 1 = "not at all valuable" to 5 = "very valuable." "A specific close contact" was perceived as the feedback source most likely to be valuable (median = 4). "A specific small group of your contacts" was considered a neutral value proposition (median = 3), and a "topical expert/professional" received a slightly lower-than-neutral rating of 2.5. Receiving feedback from "your entire social network" or from "a group of strangers" were both judged to be not at all likely to be valuable, both receiving a median rating of 1.

Contacting Others While Shopping: Current Practices

We asked respondents about their experience with contacting others while shopping: "Have you ever been shopping in a store and contacted another person about your shopping activity using your mobile phone?" 54% (107 people) reported engaging in this behavior. Mann-Whitney U tests were conducted to evaluate the statistical significance of reported differences in engaging in this activity among different demographic groups. Women were more likely to have engaged in this behavior than men (63% vs. 45%; z = 2.38, p = .017); younger people were also more likely to engage in this behavior (73% of those under 35, vs. 45% of people 35 and over; $z = 3.55$, $p < .001$), as were smartphone owners (69% of smartphone owners had done this, vs. 31% of non-smartphone owners; z = 5.21, $p < .001$).

The 107 people who reported having contacted others while shopping were then asked several follow-on questions, using a recent critical incident approach [5], in which they were asked to provide a free-text response describing the most recent time they had contacted someone while shopping. They were then asked several follow-up questions about this specific incident. Analyses and percentages in the remainder of this section refer to these 107 respondents.

Examples of the incidents participants described included:

- "*I was in Moe's Books, and called my husband to see if he wanted a particular book they had available.*"

- "*I was in Target looking for a shirt, and I couldn't decide between two. I had to call my friend to help me decide.*"

- "*I was in one of the local book/clothing stores on campus and texted my sister pictures of shirts I thought she would like.*"

- "*called mom about which dress looked better for me to wear to daughter's wedding, [at] jc penny*"

- "*I contacted my mom on which set of speakers (Logitech or BOSE) to buy at Walmart.*"

- "*I called my girlfriend from Buckle to ask if I should buy a pair of jeans.*"

The most common items involved in the described shopping situations were food (37%), followed by clothing (24%) and electronics (10%). 53% of the situations were cases where the respondent was shopping for an item for him/herself, whereas 33% were shopping for an item on behalf of the person they contacted (the remainder were shopping on behalf of a third party).

The person contacted was typically a spouse/partner (31%), a close friend (24%), parent (20%), sibling (11%), or child (8%) of the shopper; 6% responded with other relationships. Typically, the contacted person was not present on the shopping trip because they were busy with other tasks (53%); 24% were too far away to join in the shopping trip (i.e., because they lived in a different geographic region), 15% were not asked to join the shopping experience because the respondent didn't realize they would need their input ahead of time, and 8% didn't accompany the participant for other reasons.

The purpose of the contact was most often to get an opinion about a particular item (in 55% of cases), to inform the other person of an item, sale, or store they might enjoy (32%), or to ask if the person needed anything from the store (19%). In most cases, the shopper needed a response immediately (46%) or within a few minutes (42%). Most
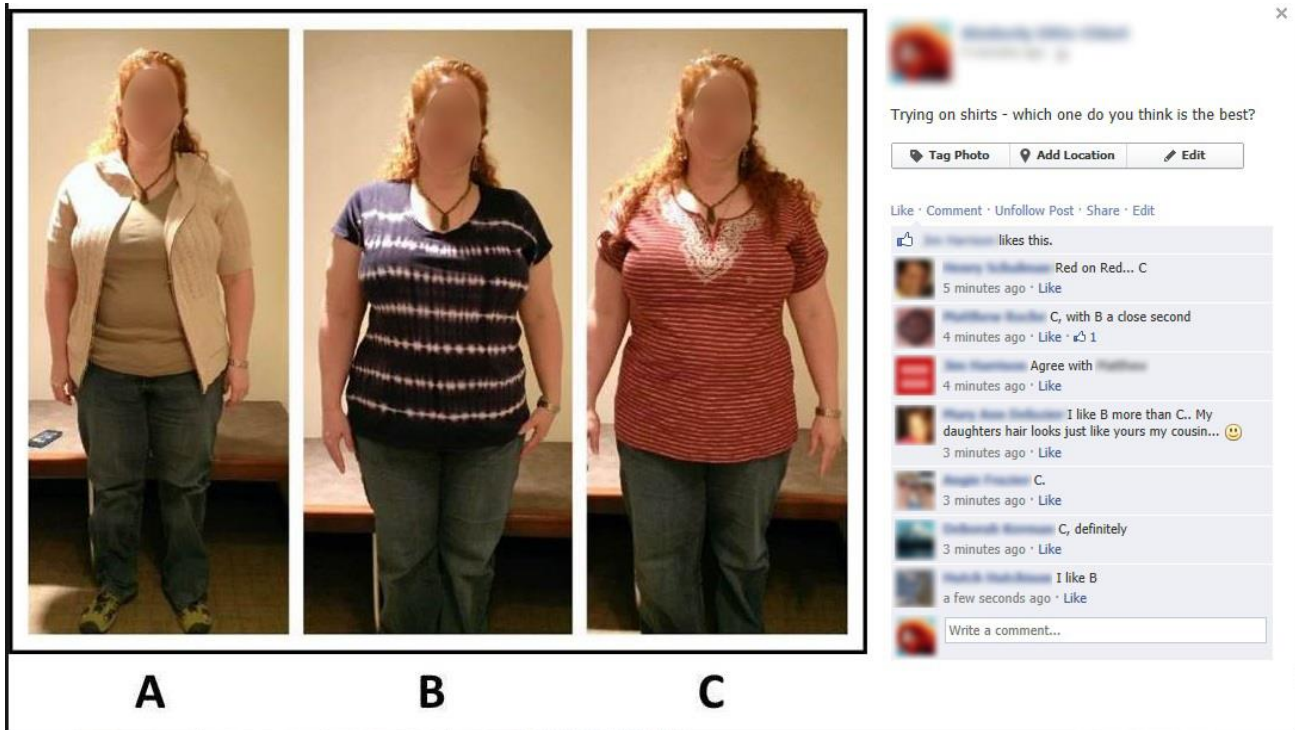
Figure 1. We generated a composite image showing the user's shopping choices, with labels ("A, B, C") to allow easy reference/voting by recipients. In this example, a participant is considering which of three shirts to purchase. Note that facial features are blurred for privacy in the paper, but were not blurred during the study itself. This screenshot shows a participant's Facebook post and the responses received. The participant added a caption to the photo: "*Trying on shirts – which one do you think is the best?*" and received opinions from seven friends within 10 minutes (with a delay of five minutes before receiving the first response). In this instance, the majority (five friends) indicate a preference for item C. Note how the third user's response builds on that of the second user: "*Agree with <name>.*"

received the desired quick replies (70% immediately, 25% within a few minutes).

Phone calls were the most common method of contact (72%), followed by MMS/SMS (47%). Some participants used multiple methods to reach out, e.g., MMSing a photo and then calling the person to discuss it. 38% of respondents shared a photo as part of this experience, and an additional 10% mentioned that they would have liked to share a photo but were unable to. Video was less popular, with only 14% having used it and another 12% indicating they did not use video, but would have liked to.

FIELD EXPERIMENT

Inspired by how common our survey revealed contacting others while shopping to be, particularly the most common scenario of seeking an opinion about an item (often with photographic support), we designed an experiment to explore the potential of emerging social platforms (social networking sites and crowd labor markets) to enhance in-store shopping. Our goal was to better understand the potential of, and differences between, targeted friendsourcing, broad friendsourcing, and crowdsourcing as enabling platforms for remote shopping advice.

We conducted the experiment in May 2013, at a Seattle-area branch of the U.S. clothing chain Eddie Bauer, which sells casual men's and women's apparel. In order to participate, participants needed to own a smartphone, have a Facebook account, and be willing to share a photo of themselves on both Facebook and Amazon's Mechanical Turk service as part of the study. Participants received a $50 Eddie Bauer gift card as a gratuity.

Participant Demographics

Participants were recruited from the local community via a recruiting service, and consisted of 14 adults[1] (6 men and 8 women), ranging in age from 20 to 55 years (mean = 32), with diverse occupations such as massage therapist, network consultant, lawyer, personal trainer, stay-at-home mother, and information technology project manager.

Participants were generally familiar with MMS/SMS as a means of communication, with 12 reporting sending basic

---

[1] Data from a fifteenth participant (P_3) was discarded due to poor network connectivity that prevented her from using the MMS feature on her phone.

text messages every day, one sending them a few times a week, and only one indicating they used text messaging rarely (less than once a month). Five participants reported that they typically send photo MMS messages every day, four others said they did so a few times a week, four a few times a month or less often, and one not at all.

All participants reported viewing Facebook to read others' posts at least a few times per week (with 9 doing so every day). All had prior experience posting photos as Facebook status messages; half reported doing so at least a few times a week, with the other half doing so more rarely. Four reported never having used their status message to ask a question of their networks, eight did so only rarely (less than once a month), and two did so often (a few times a week or every day).

10 participants (71%) were familiar with the concept of paid crowdsourcing prior to their experience in the experiment, though none had used such a service.

Methodology

When a participant arrived at the store, they were given 10 minutes to browse for merchandise to find three items they would be interested in trying on. Participants then tried on each article in a private dressing area. After trying on the first item, we instructed participants to use their own smartphone to take a photo of themselves modeling the item of clothing, in order to understand potential usability difficulties in having solo shoppers capture this type of data. One of the experimenters then used her own phone to capture an image of the participant, and also took photos of the participant modeling the subsequent two items. The experimenter than created a single composite image of the three fashion choices using the PhotoGrid app [photogrid.org], which she then manually modified to add the labels "A, B, C" below the three respective components of the image (left panel of Figure 1). The experimenter then emailed this composite image to the participant, who downloaded it onto his or her phone.

The participant next sent the composite image along with a message to one or more people by MMS, and posted the image and a (possibly different) message on their Facebook Wall (visible to all friends). Meanwhile, the experimenter posted a survey to Mechanical Turk, in a human intelligence task (HIT) that paid U.S.-based workers 25 cents to answer a four question poll giving fashion advice; the HIT had a lifespan of 10 minutes and was set to accept a maximum of 20 workers[2] (25 cents for a two-minute poll translates to $7.50 an hour, an attempt to give an ethical wage in accordance with the proposal of Kittur et al. [12]). Workers completing the HIT were redirected to a survey on

_____

[2] In one instance 21 workers completed the survey due to a race condition in which some workers began the task before others had submitted their results to Amazon.

Figure 2. Common self-portrait problems: (a) flash reflections in mirror, (b) eyes gazing at phone rather than directly ahead, and phone partially obscuring face.

surveygizmo.com showing them the composite image and asking them to (1) recommend either A, B, or C (in response to the question "Which outfit should this [man / woman] buy?", where the term "outfit" was sometimes substituted with more specific items such as "shirt" or "sunglasses" depending on what the participant chose to try on), (2) enter a brief (single-sentence) explanation explaining their recommendation, (3) specify their gender, and (4) specify their age bracket.

Meanwhile, the participant filled out a questionnaire similar to the survey about using a mobile phone while shopping described earlier in this paper. The survey was augmented with additional questions about which of the three items the participant preferred, their level of confidence in their choice, and the expected value of feedback from MMS, Facebook, and Mechanical Turk platforms. The survey took approximately 10 minutes to complete (if fewer than 10 minutes had passed, the experimenter engaged the participant in conversation to allow each medium to have 10 minutes to gather responses). 10 minutes was chosen based on the initial survey, which found that 88% of people need responses to their shopping questions either "immediately" or "within a few minutes" – we inferred from this that responses beyond a 10 minute window would therefore be of less value in many common shopping scenarios.

The participant next reviewed the responses from the three sources, by checking their phone for any MMS replies, checking their Facebook account for any comments on their post, and viewing a report generated via the Survey Gizmo service that displayed all the results from the Mechanical Turk survey, including breakdowns of the favorite item among different ages and genders of workers, and the comments supporting those choices. We recorded the number of responses from each source and the time (in minutes) to receive the first response (if any was received). Only unique people's comments were counted (i.e., if a

person sent multiple MMS replies, or made multiple Facebook posts, that only counted as a single response; Facebook "likes" were not counted as responses).

The participant then completed a final questionnaire, which asked them again to choose their favorite item, as well as to rate the value and influence of the feedback from the three sources, and to identify positive and negative aspects of their experiences with each. The entire experimental session lasted about 45 minutes. Participants were not obligated to purchase any of the items (though four chose to do so).

RESULTS

Self-Portrait
Participants used two strategies to obtain the self-portrait. 10 participants used the dressing room's mirrors to take a photo of themselves, and 4 held the phone out in front of themselves to snap a portrait. This latter method was used primarily to capture headshot-only portraits (for two men trying on sunglasses and one woman trying on hats), or head-and-torso (for one man trying on T-shirts). P_{14} noted, "*I don't do the mirror picture thing, I hate when people do that.*" Indeed, the mirror portraits suffered from many challenges, including flash artifacts in the final photo due to reflections (Figure 2a), the camera blocking part of the user's face or body in the reflected image (P_4 noted with dismay, "*it blocked out my face!*", Figure 2b), and the user's facial expression being unusual (eyes looking at the phone rather than straight ahead, as in Figure 2b, which was concerning particularly to P_6 who noted, "*normally if I take a picture of myself in an outfit, I don't smile or anything, but if it's going to go on Facebook then I want to look cute.*"). P_9 noted, "*I've never done this before, pictures of myself in the mirror.*"

Number and Speed of Responses

MMS
Seven participants sent an MMS message to exactly one person, with the rest sending to between two and six recipients. Recipients were described as close friends (7), spouses/partners (5), parents (4), siblings (4), and other relatives (2). The number of recipients was negatively correlated with time to receive an MMS reply ($r = -.2$), and the total number of replies was positively correlated with the number of recipients ($r = 0.5$).

Five of the 14 participants (36%) did not receive a reply to their text message within 10 minutes. For those who did receive MMS replies, the mean time to first response was two minutes. The average number of MMS responses received was 0.8 (rising to 1.2 when only considering people who received at least one response); nobody received MMS responses from more than two people.

Facebook
Three participants (21%) did not receive any Facebook replies within 10 minutes while another 2 participants received Facebook responses that did not offer guidance on the shopping task. These responses either sought
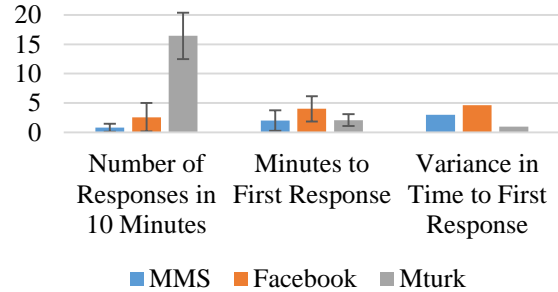

Figure 3. Relative performance of the three social feedback mechanisms used in our field experiment (error bars reflect standard deviation). Crowdsourcing via Mechanical Turk quickly and consistently provided the most responses. MMS performed more quickly than Facebook on average, but was less consistent in producing a response. Time to first response reflects only those participants who received responses in a given medium.

clarification or were humorous (e.g., "*no swimwear option?*" (P_{10}), "*where are you going?*" (P_{12})).

An average of 2.6 Facebook responses were received (3.3 among those who received at least one answer, min = 1, max = 8), with an average of 4 minutes to receive the first response. Figure 1 depicts an example of a Facebook post and the responses obtained during the experiment.

Participants had between 179 and 1559 Facebook friends (median 340, mean 512). As expected, larger network sizes increased Facebook's responsiveness, with the number of responses positively correlated with network size ($r = 0.2$) and time to first response negatively correlated ($r = -0.6$). Factors beyond network size may also have had an impact, such as time of day – for instance, P_8 had immigrated to the United States from India a few years prior, and noted that over half of his Facebook friends still lived in India; due to time zone differences, those friends were likely asleep and therefore not available to see his post (he received no replies). Facebook use habits also impacted performance, with participants' self-reported frequency of posting a status update correlating positively with the number of responses received ($r = .4$) and negatively with time to first response ($r = -.5$).

Crowdsourcing
All participants received feedback from Mechanical Turk within the allotted time period. An average of 16.4 Mechanical Turk workers provided feedback within 10 minutes (min = 8, max = 21), with an average of 2.1 minutes to the first completed response (the four-item survey took between one and two minutes to complete).

Comparisons
Figure 3 illustrates the responsiveness of all three communications platforms used. A one-way repeated measures ANOVA found a significant difference in the

number of responses received from each source ($F_{2,12}$ = 99.9, $p < .001$); follow-up pairwise comparisons between each feedback source were all significant ($p \leq .01$) with Mechanical Turk receiving the most responses (a mean of 16.4), followed by Facebook (2.6), and MMS (0.8).

Time to first response was not significantly different across feedback sources when considering only those cases where feedback was received from all three sources, $F(2,5) = 1.28$, $p = .36$. However, using 11 minutes as a generous estimate of the time to first response for people who did not receive a response from a given medium within 10 minutes, a one-way repeated measures ANOVA found a significant difference in the time to first response: $F(2, 12) = 5.34$, $p = .02$. Follow-up pairwise comparisons between time to first response for each feedback source showed no significant difference in response time between MMS and Facebook ($p = .84$), but a significant speed advantage for Mechanical Turk as compared to MMS ($p = .037$) and Facebook ($p = .008$). Furthermore, the speed of response from Mechanical Turk was more consistent, having the least variance in time to first response (1.0), followed by MMS (3.0) and then Facebook (4.6).

Response Quality
MMS replies tended to be quite short (which is understandable, due to the character limits imposed on MMS messages, the difficulty of typing on smartphones, and the social conventions surrounding texting). Examples of typical, brief MMS replies included: "*B*"; "*C. The watch.*"; "*The first one.*" Only two participants received detailed replies via MMS: "*C seems to stand out, A, too, and I could see you choose B (I know, no help)*"; "*Yes, black suits you. And I like the first one also. A or C. A seems very summery.*" The average length of MMS responses was 27.9 characters.

Facebook responses included a roughly equal mix of very brief (e.g., "*A,*" "*I like B,*" "*C, definitely*"), and those elaborating on the reasoning behind the choice ("*... I always love more color... C is my choice*"; "*I like either A or B because they would work better through all the seasons.*"; "*C! Lights up your face!*"). Responses were generally positive, though three male participants (P1, P9, P12) and one female (P10) each received a response that employed sarcastic humor ("*That hat is freaking terrible. Burn it.*"; "*I would kiss the guy in A*"; "*No swimwear option?*"). Facebook responses averaged 23.4 characters

All of the crowd-workers were required to complete a free-response question that asked them to "explain the reasoning behind your recommendation" (of item A, B, or C in the prior survey question). Crowdsourced responses averaged 63.1 characters.

A repeated-measures ANOVA indicates a significant difference in response length: $F(2,5) = 25.7$, $p = .002$; follow-up pairwise comparisons indicate no significant difference in length of response from MMS and Facebook ($p = .61$), but the crowdsourced responses were significantly longer than either MMS responses ($p = .037$) or Facebook responses ($p = .001$). Longer length has been found to be a positive indicator of answer quality in other social media, such as online Q&A forums [7].

We counted the number of generic, low-quality responses from Mechanical Turk (e.g., responses that simply stated that one choice looked best, without offering any specific rationale, such as "*looks best*" or "*It's the most pretty*"). The proportion of such responses ranged from 0% to 26%, averaging 10.6% of all responses. The overwhelming majority of crowd workers offered specific tidbits of thoughtful advice, such as: "*I think the floral pattern is very pretty and adds a bit of flair*"; "*It's sophisticated but casual, and fits her the best.*"; "*I like that it is plain but a good fit.*"; "*It's simple and clean and well [sic] work for any occasion [sic].*"; "*The colors go good with her hair.*"

Unlike MMS and Facebook feedback, which only offered positive comments (with the exception of the few humorous responses), the responses from the crowd workers also gave feedback about which items were not flattering, and why. For example, "*The green shirt does not look good at all, doesn't go well with her face and eyes...*"; "*The white V-neck is kinda cheesy with the flower print.*"; "*...A makes his arms look super short*"; "*the other colors are boring*"; "*It is more for his age.*" Many participants valued this honesty (P4 noted this contrast with Facebook – "*sometimes your friends lie to you* [by not telling you what looked bad]").

Participants found the ability to see breakdowns of crowd workers' votes by age and gender "*very cool... very interesting*" (P7). For example, male participant P12 found it revealing that men's votes were distributed equally across the three shirts, whereas none of the women liked the shirt that was less tight-fitting, and people aged 35 and over preferred a more conservative, collared shirt, whereas younger people preferred the other options.

For the 11 participants who received responses expressing a fashion preference from at least one of the personalized social sources (MMS or Facebook), in 9 cases (82%) the majority choice of the Mechanical Turk workers agreed with the majority choice from at least one of the personal social sources (in the 7 cases where all three sources returned an opinion, the MMS and Facebook majority opinion matched for 4 (57%) of the cases, and the Turk opinion agreed in all four of these cases). The majority recommendation of the Mechanical Turk workers was the only one to be significantly correlated with participants' own final item preference (Pearson's R = .53, p = .05).

Impact on Participants' Choice

Recommendations Match Participant's Initial Choice
For five participants (P5, P6, P7, P11, P13) their initial favorite item matched the recommendations from at least one of the feedback sources and they maintained this choice after receiving all of the feedback. Two participants (P6 and P13),
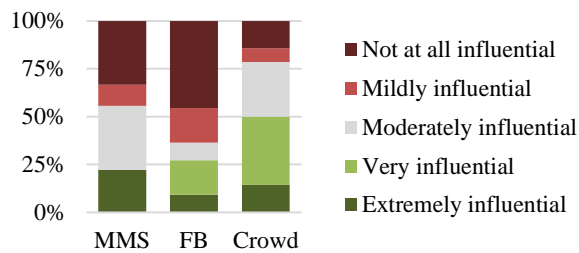
Figure 4. Participants reported that the feedback crowdsourced from strangers was influential.

however, did not align with the majority feedback. P_6 commented that "*I usually end up going with my own opinion*" and P_{13} emphasized her personal preference, "*I like the wider spacing on the stripes.*" In all cases, these participants were initially very confident with their choice (4 or 5 ratings on a five-point scale) and their confidence either stayed the same, or increased.

Participant P_4 was interesting because the recommended feedback from all three sources matched her initial choice, but after the feedback, she changed her choice. When asked what she was looking for from the feedback sources she commented, "*I wouldn't take anyone's feedback.*" Initially, she was very confident in her choice (5 rating) but after she changed her choice (against the recommendations) her confidence in her new choice was low (2 rating).

Recommendations Do Not Match Participant's Initial Choice
For four participants (P_2, P_9, P_{12}, P_{15}), their favorite items did not match the recommendations from the feedback sources, but they maintained their initial choices. Their comments indicated that they primarily stayed with their choice because of personal preference, but did consider the recommendations ("*I still like the white, but the comment that it looks like a undershirt may have some merit. Most women say that A looks better, so I would strongly consider A if B was not available*" P_{15}, "*the age group/gender that I would be most interested in looking attractive to seemed drawn to B or C, and personal preference combined with the crowdsourcing feedback convinced me to choose B*" P_9). P_9 and P_{12} were very confident in their initial choice (ratings of 4 or 5), but after getting feedback P_9's confidence went down and P_{12}'s stayed the same. P_2 and P_{15} were not confident both before and after the feedback (ratings of 1 or 2) and P_2 commented that she didn't like any of the items.

For four participants (P_1, P_8, P_{10}, P_{14}), their initial favorite item was not recommended by any of the feedback sources, and all these participants subsequently changed their choice to match the majority recommendations received. In all four cases, their new choice matched the crowd workers' modal recommendation. Three of the participants (P_1, P_8, P_{14}) were initially very confident with their choices (4 or 5 ratings) but P_{10} was not confident with her choice (1 rating). After the feedback, all four participants were very comfortable

with their new choice (4 or 5 ratings). Three of the participants explicitly commented that the crowdsourced feedback caused them to change their rating ("*The crowdsource* [sic] *& MMS responses were different than what I picked and I will go with what looks best to everyone*" P_{14}, and "*The watch seemed to get a lot of good feedback that was well reasoned via crowdsourcing. I find that surprisingly compelling*" P_1).

Usefulness of the Different Feedback Sources
When asked what kind of feedback would be most valuable in helping them decide among the items, eight participants mentioned feedback about whether the items looked good or not ("*opinions of whether they look good or not, fashionable, etc.*" P_5, "*what looks best*" P_{14}) and four participants mentioned feedback from a friend, boyfriend, or girlfriend ("*would want one of my girlfriends to give me a yay or nay*" P_2).

After receiving their feedback, participants ranked how influential the feedback from each source was in determining their preferred choice (on a five-point scale from "not at all influential" to "extremely influential," Figure 4). Participants rated the feedback from the crowdsourcing platform (median = 3.5) as more influential than both MMS (3) and Facebook (2), though this trend is not a statistically significant difference according to the results of a Friedman test, $\chi^2(2, N=7) = 3.7$, p = .156 (note that the test excludes cases where participants did not receive feedback from all three sources).

Before receiving feedback, the participants rated how useful they felt each source would be (on a five-point scale from "not at all useful" to "extremely useful") and after receiving feedback they rated the actual usefulness of each source (for those sources from which they received feedback). For MMS, the median rating was 3 ("moderately influential") at both the start and end of the study. A Wilcoxon test indicates that there was no significant difference in participants' opinions of the expected utility of MMS feedback at the beginning of the study with their opinions of its actual utility at the end of the study, $z = 1.5$, $p = 1.0$. A similar trend was found for Facebook, which also had a median rating of 3 in both instances, $z = -1.12$, $p = .27$. (Note that these median ratings likely reflect an overestimate of the perceived utility of MMS and Facebook, as only participants who actually received feedback from these sources rated them – participants who received no feedback at all presumably received a much lower utility from such sources.) The results for the crowdsourced condition however, were significantly different, with the median rating for the perceived utility rising from 3 ("moderately useful") before the experience to 4 ("very useful") after having received feedback, $z = 2.49$, $p = .013$.

Advantages & Disadvantages of the Feedback Sources

MMS

When asked what they liked best about MMS, the main reasons given were the ability to ask specific people (P_1, P_5, P_6, P_9, P_{14}, P_{15}) ("*trusted individual, so the opinion matters more*" P_5, "*exactly the people I want to hear from*" P_6) and that it was quick and easy (P_2, P_7, P_9, P_{13}) ("*a convenient way for me to contact people and receive their responses*" P_7). The main disadvantages expressed for MMS was the potential delay in response time (P_1, P_6, P_7, P_8, P_9) ("*No idea if you'll get feedback from any given person within the timeframe you need*" P_1). Other reasons included not being able to get a large number of opinions (P_{15}), and getting terse responses (P_5).

Facebook

When asked what they liked best about Facebook, the main reasons given were that it was a close community of friends (P_5, P_8, P_9, P_{11}) ("*feedback from my closer community of friends*" P_8) and that they could get feedback from lots of people (P_6, P_9, P_{14}, P_{15}) ("*feedback from a large group of people*" P_{15}). Others commented on the benefit of getting insightful, interesting, or humorous comments (P_1, P_2, P_{10}). The main disadvantages expressed were replies that weren't very useful (P_1, P_4, P_5, P_7, P_{12}, P_{13}), and that it can take time to get responses (P_9, P_{11}, P_{15}) ("*it could take too much time*" P_9, "*they would have to be online at the time and view my picture which they may not see because their newsfeed may be too large and my status and pics may be hidden from them*" P_{15}).

Several participants expressed varying degrees of discomfort with posting the photo-question to Facebook (note that during the study's recruiting phase, participants were forewarned that they would be asked to post a photo to both Facebook and Mechanical Turk if they chose to participate). P_3 (the participant whose data was not used due to a network connectivity issue that prevented MMS messages from being sent) was so uncomfortable with posting to her Facebook Wall that she instead constructed a private Facebook message and sent the photo to only a few close friends. P_4 deleted her Facebook post immediately after the 10 minute response-gathering period was complete. P_8 mentioned that he intended to immediately delete his post (explaining that his manager was on Facebook). P_6 noted that she typically doesn't post photos (although she used to), noting that, "*I decided to be more selective in the last year*" after realizing how large and diverse the audience of her Facebook friends was. P_9 phrased the caption on his photo in a facetious manner ("*Hello friends, I need to know which of these shirts makes me look most attractive/radiant.*"), and when asked about this revealed that the humorous phrasing was due to his discomfort, "*it's not something I would normally do, I don't know when the last time is I actually posted a picture, it's been a while.*" P_{10} commented in the post-study questionnaire that a drawback of using Facebook was that "*everybody sees it.*"

Crowdsourcing

When asked what they liked best about crowdsourcing, the main comments given (by seven participants) were that the feedback came from strangers who could be objective and give a wide variety of responses ("*different people's opinions*" P_{12}, "*representation of the sorts of opinions people on the street would have*" P_7, "*honest opinion*" P_{15}). Five participants liked that they could get many responses quickly ("*lot of feedback in a very short amount of time*" P_9), and four favorably noted that they could get the breakdown of age and gender ("*Breakdown by age and gender. I can make decisions based on what demographics things do well with*" P_1). The main disadvantages expressed were the fact that the respondent do not know them or their style (three participants) ("*Lack of familiarity with what my general style would be like*" P_1, "*Hard to trust opinion of people I don't know*" P_6) and that sometimes there were too many differing opinions (two participants) ("*Opinions were all over the map*" P_5). While some people appreciated the feedback they received ("*responses were very positive and many of the people gave very specific reasons for their choice*" P_{11}), others were uncomfortable with the bluntness of some of the comments. For example, after a crowd worker wrote, "*she [P_{14}] is a bigger woman and the bigger hat kinda hides the bigness,*" P_{14} expressed that she disliked "*[the] comment about my size.*"

DISCUSSION

Our initial survey revealed that many people engage in remote shopping advice experiences via mobile phones, typically via traditional voice-based phone calls, though occasionally employing multi-media including text, photos, and video. These experiences always involved known contacts (friends and family), and survey respondents did not think that social shopping feedback from strangers would be valuable (except perhaps topical experts). However, when exposed to rich, multimedia remote shopping advice experiences in our field experiment (posting photos via MMS, Facebook, and Mechanical Turk), participants reacted favorably, finding feedback from all sources useful and influential. Participants were often surprised by the usefulness of feedback from strangers; their attitudes about this novel social experience changed significantly during the course of the session.

Friends vs. Strangers

The main strengths of receiving feedback from crowds of strangers, rather than from friends (via MMS or Facebook), included:

Independent judgments: Crowd workers did not see others' votes, and were not influenced by them (a prerequisite to effectively harnessing the "wisdom of crowds" [23]); in contrast, Facebook users were sometimes perceived as amplifying the opinions of those who replied earliest (in some cases this echo-chamber effect was explicit, e.g. the third response in Figure 1, which notes "Agree with [name of prior commenter]").

Honesty: While personal contacts told users what items they recommended, the crowd workers also told users what items they *didn't* recommend, and why. Most users found these critical opinions refreshing, though a few were insulted (e.g., the woman whose size was referred to in a blunt manner). Granovetter [6] identified the benefit of "weak ties" for providing novel information; our findings indicate that "non-ties" also provide novel information (critical/negative feedback) not offered by either their strong or weak personal ties.

Speed and consistency: Both personal contacts (via MMS and Facebook) and strangers (via Mechanical Turk) responded within a few minutes of receiving the user's inquiry. Our initial survey study found that feedback within a few minutes would be satisfactory to most people seeking remote shopping advice, suggesting that this level of latency is acceptable for creating a working system. Crowd workers' response latency (and likelihood) was much more consistent than personal contacts, though, who were not always immediately available. However, the performance of Facebook was much faster than in studies from only a few years ago (e.g., [25], which found latencies of nearly an hour to be typical). We attribute this improvement in Facebook's performance to several causes: the increasing penetration of Facebook (since larger networks result in faster response times), the increasing penetration of smartphones (since these allow people to be connected to social media in a larger number of settings, and therefore available to respond to posts), and the use of a photo, rather than a text-only post (although Facebook's Newsfeed algorithm is a proprietary secret, informal observations suggest that posts containing photos seem to be given more prominence).

The main drawbacks of using crowds rather than personal contacts were:

Context: As found in prior studies of social network question asking (e.g., [17]), users appreciate that their personal contacts are often aware of relevant context (in this case, a user's personal style). Additionally, other types of context (such as the price of the items being considered) might be relevant to provide, regardless of whether the answerers are personal contacts or strangers.

Cost: In this experiment, we absorbed the cost of crowdsourcing, but presumably this would be paid by end users in a deployed system. We paid a relatively high cost (25 cents per response) in order to both provide a fair wage [12] and encourage speedy responses. Other mechanisms could be used to facilitate fast replies (e.g., techniques proposed by Bernstein et al. [1] and Bigham et al. [2] for real-time crowdsourcing), but the cost of crowd opinions is still likely to be higher than the free price of responses from personal contacts. However, such responses come with a social cost [3]; for frequent inquiries, many users may prefer a financial cost rather than a social one. Discovering how much participants are willing to pay for such a service

in practice is an open question; it may be that participants' enthusiasm for the service would wane if they bore the cost of crowdsourcing rather than our research team.

Privacy did not appear to be a drawback of crowdsourcing, though participants did express concern about sharing images of themselves to their personal networks on Facebook (though perhaps that is a an artifact of the type of participant willing to volunteer for our study; it is possible that people who would have had concerns about sharing images of themselves on Mechanical Turk chose not to participate). Blurring or other techniques such as [15] could be used to mitigate potential privacy issues on Mechanical Turk, although facial features, skin tone, hair, etc. were a factor that influenced many crowd workers' choices, as revealed through their comments (e.g., "*doesn't go well with her face and eyes*"; "*The jacket matches well with her skin complexion and hair*"). The EmailValet project [13], where paid crowd workers triaged a users' email inboxes, is another example of end-users finding value in sharing potentially private information with paid crowd workers. Assessing the risks, benefits, and tradeoffs involved in sharing various types of information with members of the crowd is an important area for additional research. The use of mechanisms like Facebook's Lists or Google Plus's Circles to restrict sharing to a subset of one's social network is a possible mechanism for mitigating users' concerns regarding sharing in that medium, where issues regarding curating a particular type of public image drive many users' decisions about what to post [28].

Towards Remote Shopping Advice Systems

Our survey's findings indicate that seeking input from remote people while shopping is a relatively commonplace occurrence, but that most people currently rely on simple voice or text-based interactions to accomplish this. Our experiment demonstrated that users found value in using richer media (photos) as well as using emerging social platforms (social networking sites and crowd labor markets) to meet these needs, and that such platforms' performance characteristics (particularly Mechanical Turk) were generally suitable for such interactions.

Based on these findings, we suspect that consumers would find value in a smartphone app designed specifically to support seeking remote shopping advice. Our results suggest that key capabilities of (and challenges that must be overcome by) such an app would include:

Image capture: Facilitating a user's ability to capture an image, particularly for clothing, which requires self-portraits, is a challenge. One approach, such as that used by the app "Headshot" [aka.ms/headshot] is to provide visual feedback to help the user better position the camera. Using video and perhaps letting crowd workers choose the best frame for inclusion in an image [1] may be an alternate approach. Crowd labor or automatic techniques could also be used to automate the manual image compositing/labeling we did for our study (we recommend using a single,

composite image, due to users' reluctance to bombard their social networks with too many posts). Beyond technical issues, capturing such images or video may be challenging due to evolving societal norms about the use of such technologies – other patrons in dressing rooms may have privacy concerns about being inadvertently included, or shop owners may assume patrons are recording images of merchandise as a reminder to later seek better deals in online shops.

Audience targeting: Providing the ability for users to target their query to one or more audience types would enable users to (if desired) harness both the consistent speed and "blunt" responses from sources like Mechanical Turk as well as the personalized and trusted responses from friends and family. Such a system could show users predictions of the likely response time from each source, based on factors like the price they are willing to pay for crowd labor, the time of day, and the size of their online social networks. Of course, users' choices of which platforms to employ might be influenced by characteristics beyond answer speed and type – privacy preferences, and/or differential preferences for the "informational" versus "social" aspects of the remote shopping advice experience may also influence their selection of medium. Users choosing to engage crowd laborers could potentially specify worker characteristics that were relevant to their task, such as age, gender, geographic region, expertise with certain types of products, personal taste profiles, etc.

Decision support: Finally, an ideal remote shopping advice application would provide interactive support (perhaps through information visualization techniques) to allow users to explore, compare, and contrast feedback from different audiences (participants in our study found the differences and similarities between personal contacts' and strangers' recommendations informative, as well as differences between sub-audiences, such as male versus female workers). Such an interface might also allow users to factor in other sources of information, such as online reviews and pricing.

CONCLUSION

In this paper, we presented a formal study of *remote shopping advice*, in which people use their mobile phone to contact others in support of in-store shopping tasks. We presented results from our multi-method study, comprising (1) survey results about users' status quo practices, and (2) experimental results comparing and contrasting the use of different social technologies to support seeking remote shopping advice (MMS for small groups of close ties, Facebook for extended personal networks, and Mechanical Turk for paid crowd workers). We also presented design guidelines for remote shopping advice systems synthesized from the findings of our survey and experiment. Our key finding was that the crowdsourced feedback was surprisingly useful to our participants – participants were influenced by the high-quality and honest nature of crowd

workers' comments, despite the lack of context and potential for privacy concerns. In fact, they were more reticent to share their photos on Facebook than Mechanical Turk. Our findings illustrate how smart mobile devices connected to powerful communications platforms have the potential to transform even mundane daily tasks into experiences that can inform, connect, and delight.

REFERENCES

1. Bernstein, M., Brandt, J., Miller, R., and Karger, D. Crowds in Two Seconds: Enabling Realtime Crowd-Powered Interfaces. *Proceedings of UIST 2011*.

2. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. VizWiz: Nearly real-time answers to visual questions. *Proceedings of UIST 2010*.

3. Brady, E., Zhong, Y., Morris, M.R., and Bigham, J. Investigating the Appropriateness of Social Network Question Asking as a Resource for Blind Users. *Proceedings of CSCW 2013*.

4. Burton, M.A., Brady, E., Brewer, R., Neylan, C., Bigham, J.P., and Hurst, A. Crowdsourcing subjective fashion advice using VizWiz: Challenges and opportunities. *Proceedings of ASSETS 2012*.

5. Flanagan,J. The critical incident technique. *Psychological Bulletin*, 51:327-358, 1954.

6. Granovetter, M. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6), 1973, 1360-1380.

7. Harper, F.M., Raban, D., Rafaeli, S., and Konstan, J.A. Predictors of Answer Quality in Online Q&A Sites. *Proceedings of CHI 2008*.

8. Hillman, S., Neustaedter, C., Pang, C., and Oduor, E. "Shared Joy is Double Joy": The Social Practices of User Networks within Group Shopping Sites. *Proceedings of CHI 2013*.

9. Hillman, S., Neustaedter, C., Bowes, J., and Antle, A. Soft Trust and mCommerce Shopping Behaviors. *Proceedings of MobileHCI 2012*.

10. Jeong, J-W., Morris, M.R., Teevan, J., and Liebling, D. A Crowd-Powered Socially Embedded Search Engine. *Proceedings of ICWSM 2013*.

11. Jung, Y., Gray, R., Lampe, C., and Ellison, N.B. Favors from Facebook Friends: Unpacking Dimensions of Social Capital. *Proceedings of CHI 2013*.

12. Kittur, A., Nickerson, J.V., Bernstein, M.S., Gerber, E.M., Shaw, A., Zimmerman, J., Lease, M., and Horton, J.J. The Future of Crowd Work. *Proceedings of CSCW 2013*.

13. Kokkalis, N., Köhn, T., Pfeiffer, C., Chornyi, D., Bernstein, M.S., and Klemmer, S.R. EmailValet: Managing e-mail overload through private, accountable crowdsourcing. *Proceedings of CSCW 2013*.

14. Lampe, C., Vitak, J., Gray, R., and Ellison, N.B. Perceptions of Facebook's Value as an Information Source. *Proceedings of CHI 2012*.

15. Lasecki, W.S., Song, Y., Kautz, H., and Bigham, J.P. Real-Time Crowd Labeling for Deployable Activity Recognition. *Proceedings of CSCW 2013*.

16. Miller, D., Jackson, P., Thrift, N., Holbrook, B., and Rowlands, M. Shopping, Place and Identity. *Routledge Publishing*, New York, NY, 1998.

17. Morris, M.R., Teevan, J., and Panovich, K. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. *Proceedings of CHI 2010*.

18. Panovich, K., Miller, R., and Karger, D. Tie Strength in Question & Answer on Social Network Sites. *Proceedings of CSCW 2012*.

19. Pentina, I., Prybutok, V., and Zhang, X. The Role of Virtual Communities as Shopping Reference Groups. *J. Electronic Commerce*, 2008.

20. Rainie, L. Smartphone Ownership Update: September 2012. *Pew Internet & American Life Project*.

21. Smith, A. In-store Mobile Commerce During the 2012 Holiday Shopping Season. *Pew Internet & American Life Project*. January 31, 2013.

22. Strauss, M. Sears shoppers make fashion link with Skype. *The Globe and Mail*, April 21, 2011.

23. Surowiecki, J. The Wisdom of Crowds. 2005.

24. Tauber, E., Why Do People Shop? Marketing Notes and Communications, *J. Marketing*, Vol. 36, (1972).

25. Teevan, J., Morris, M.R., and Panovich, K. Factors Affecting Response Quantity, Quality, and Speed for Questions Asked via Social Network Status Messages. *Proceedings of ICWSM 2011*.

26. Tohidi, M. and Warr, A. The Bigger Picture: The Use of Mobile Phones in Shopping. *Proceedings of Interact 2013*.

27. Tsujita, H., Tsukada, K., Kambara, K., and Sijo, I. Complete Fashion Coordinator: A support system for capturing and selecting daily clothes with social network. *Proceedings of AVI 2010*.

28. Zhao, X., Salehi, N., Naranjit, S., Alwaalan, S., Voida, S., and Cosley, D. The Many Faces of Facebook: Experiencing Social Media as Performance, Exhibition, and Personal Archive. *Proceedings of CHI 2013*.