Attractive Phones Don't Have To Work Better: Independent Effects of Attractiveness, Effectiveness, and Efficiency on Perceived Usability

Jeffrey M. Quinn
Sprint Nextel
6220 Sprint Parkway
Overland Park, KS
jeffq1126@yahoo.com

Tuan Q. Tran
Sprint Nextel
6220 Sprint Parkway
Overland Park, KS
Tuan.Tran@sprint.com

ABSTRACT
Participants sometimes rate products high in usability despite experiencing obvious usability problems (low effectiveness or efficiency). Is it possible that this occurs because high product attractiveness compensates for low effectiveness/efficiency? Previous research has not investigated the interplay between attractiveness, effectiveness, and efficiency to determine whether attractiveness accounts for additional variance in usability ratings beyond that which is explained by effectiveness and efficiency. The present research provides the first test of this idea. Using data from usability testing, we demonstrate that attractiveness, effectiveness, and efficiency each has an independent influence on usability ratings and, in the present research, attractiveness had the largest impact. We report results of quantitative analyses that suggest multiple mechanisms could be responsible for the relationship between attractiveness and usability.

Author Keywords
Mobile phone, attractiveness, aesthetics, system usability scale, SUS, need for cognition.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User interfaces – Evaluation/methodology; Benchmarking.

General Terms
Experimentation, Measurement

INTRODUCTION
In our work testing the usability of mobile phones we find that, typically, participants who can use a phone effectively and efficiently to complete common phone tasks tend to rate the phone high in usability whereas participants who cannot tend to rate the phone low in usability [28]. Occasionally, however, participants' ratings of a phone's usability are different than might be expected based on their performance during the testing session [13]. For instance, participants sometimes rate a phone high in usability despite experiencing obvious usability problems [2, p. 119].

What is it about a phone that could lead consumers to perceive that the phone is easy to use despite experiencing difficulty performing common tasks? Results of several recent studies suggest that the attractiveness of a product's appearance can influence perceptions of usability [e.g., 10, 32, 33]. In these studies participants rated attractive products higher in usability than unattractive products. However, the finding of a correlation between attractiveness and perceived usability does not necessarily prove that attractiveness can compensate for ineffective or inefficient task performance. Rather, it is necessary to test the *simultaneous* influence of these variables (attractiveness, effectiveness, and efficiency) on perceived usability. Such a multivariate approach would show whether attractiveness has a unique effect on perceived usability that is independent of the effects of actual task performance (i.e., effectiveness, efficiency). If it does, then it is possible that an attractive phone could be rated high in usability regardless of (low or high) effectiveness and efficiency. Alternatively, attractiveness might not have an independent influence on perceived usability. For example, it could be that the attractiveness-usability relationship varies depending on task performance outcomes (e.g., attractiveness only influences usability if the user achieves some minimal threshold of effectiveness) or that measures of attractiveness and task performance themselves are so closely associated that attractiveness has no influence on usability after controlling for effectiveness and efficiency (as might be expected if "attractive things work better" [24, p. 17; see also 10, 33]).

In the present research we tested attractiveness, effectiveness, and efficiency as simultaneous predictors of participants' ratings of a mobile phone's usability. Doing so allowed us to determine whether attractiveness had a unique effect on perceived usability that was independent of the effects of effectiveness and efficiency. In addition, it

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

allowed us to observe whether the magnitude of the attractiveness effect was large enough to counteract the impact of ineffective/inefficient performance. Our findings build upon and integrate two areas of usability research investigating the relationship between objective task performance and perceived usability and the relationship between product attractiveness and perceived usability.

Task Performance Influences Perceived Usability

Often, participants in a usability test complete a scale measure containing items such as, "I thought this system was easy to use" [4], after performing a set of tasks with a product. A basic assumption underlying this paradigm is that people come to know how usable a product is as a result of actual use and, thus, their scale ratings of usability should reflect the effectiveness and efficiency with which they used the product in the testing session. Results of empirical studies generally are consistent with this assumption. For example, in a quantitative synthesis of 73 studies, Hornbæk and Law [15] observed statistically significant correlations between participants' subjective ratings of usability and measures of task performance such that high effectiveness and efficiency were associated with high perceived usability. In another review of published studies, Nielsen and Levy [22] reported a similar relationship between subjective measures of usability (i.e., scale ratings, preferences) and objective measures of effective and efficient performance.

Results of both Hornbæk and Law [15] and Nielsen and Levy [22] also highlight the fact that effectiveness and efficiency are not the *only* factors that influence participants' perceptions of product usability. Hornbæk and Law showed that relationships between usability ratings and both effectiveness and efficiency tended to be small in magnitude, indicating that performance outcomes did not completely explain why participants rated a product's usability the way that they did. Nielsen and Levy reported that objective performance accounted for 21% of the variability in participant's subjective ratings [see also 28]. Thus, a great deal of the variability in these ratings was associated with factors unrelated to the effectiveness and efficiency of performance.

It is interesting to note that Nielsen and Levy [22] speculated (but did not test) that "factors such as graphic design quality may account for part of the remaining variance" unexplained by task performance (p. 73). To the extent that graphic design quality is similar to product attractiveness, this idea is equivalent to the hypothesis investigated in the present research, i.e., that attractiveness shapes perceptions of usability controlling for effectiveness and efficiency of task performance. To our knowledge, the present research provides the first test of this idea.

Product Attractiveness Influences Perceived Usability

Evidence from several studies indicates that a relationship exists between product attractiveness and perceived usability. For example, in a study by Chawda, Craft, Cairns, Rüger, and Heech [10], participants performed a search task then rated the attractiveness and usability of the search tool. Participants' ratings were significantly correlated such that usability was higher for search tools participants perceived as attractive versus unattractive [see similar results in 5, 19, 32, 33]. Results such as these suggest that when participants rate a product's usability, one factor that influences their ratings is product attractiveness.

In the published studies, researchers typically have investigated the simple bivariate relationship (e.g., correlation) between attractiveness and usability. Only Ben-Bassat, Meyer, and Tractinsky [5] have tested the influence of attractiveness controlling for the effects of another variable. These authors manipulated the aesthetic qualities (i.e., attractiveness) and the efficiency (i.e., number of steps required to enter data) of an address book application and instructed participants to enter names and phone numbers into four versions of the address book characterized by high efficiency/high aesthetics, high efficiency/low aesthetics, medium efficiency/high aesthetics, and medium efficiency/low aesthetics. After entering data, participants rated the usability of each version. Results of an analysis of aesthetics and efficiency as simultaneous predictors indicated that both variables were significantly associated with perceived usability. Thus, the aesthetic qualities of the address book shaped usability ratings controlling for the effects of high versus medium efficiency.

In contrast to the results of Ben-Bassat et al., other research suggests that the effect on perceived usability of attractiveness might not be independent of the effects of task performance. Norman [24] offered an explanation of the attractiveness-usability relationship that was based on the notion that attractive products actually work better (e.g., more effectively or efficiently) than unattractive products. In explaining this relationship, Norman cited two lines of research: (1) Norman and colleagues' own theoretical work describing the ways that objects (e.g., products) shape users' emotional experiences (i.e., by eliciting visceral, behavioral, and reflective affective responses) [25], and (2) research by Isen and others [see 16 for a review], demonstrating that positive affect leads to creative thought and improved performance. Combining these two ideas, Norman [24] proposed that "attractive things make people feel good, which in turn makes them think more creatively. How does this make something easier to use? Simple, by making it easier to find solutions to the problems they encounter" (p. 19). By this account, attractiveness and performance are related because, to some extent, attractiveness plays a causal role in bringing about improved performance. Depending upon how strong the attractiveness-performance relationship is, it is possible that the effects of attractiveness, effectiveness, and efficiency might overlap so much that attractiveness does not account for *unique* variance in perceived usability beyond that which is explained by effectiveness and efficiency.

The Present Research

In this paper we report results of a quantitative analysis of data from lab-based usability testing in which participants used a mobile phone to perform several tasks then rated the phone's usability and attractiveness. We performed analyses using attractiveness, effectiveness, and efficiency as simultaneous predictors of participant-rated usability. Our primary goal was to better understand why users rated the phone's usability the way that they did. Based on past research, we anticipated that the effectiveness and efficiency with which users completed tasks would explain some of the variability in usability ratings. We aimed to determine whether phone attractiveness (as rated by the user) explained additional variance in usability ratings above and beyond the variance explained by effectiveness and efficiency alone.

As a secondary and more exploratory goal, we performed additional detailed analyses of the relationships between attractiveness and other variables assessed in this work. With these analyses, we aimed to provide initial results to stimulate thought and future research on possible mechanisms underlying the attractiveness-usability relationship. Specifically, we examined whether the data in the present research were consistent with the theoretical framework described by Norman [24]. We focused on Norman's framework for several reasons: it is the most clearly defined explanation of the attractiveness-usability relationship in the published literature, it is supported by much empirical work [e.g., 16], and the present research happened to include many of the variables that comprise the mechanism Norman described. As stated previously, Norman's framework specifies a process through which attractive products elicit positive affect, which enhances creative thought and makes it easier for users to find solutions to usability problems, thus, leading to improved task performance and greater perceived usability [24; see also 16, 25]. Since the present research included measures of attractiveness, task performance, and perceived usability, it was possible to test whether the patterns of correlations among these variables in our work fit with Norman's mechanism, in which attractive things work better than unattractive things. In addition, we were able to test the possibility that effortful thought mediates the relationship between attractiveness and task performance. We did so using data we collected from participants on the personality trait, *need for cognition* (NC) [8], which measures the extent to which a person enjoys and engages in effortful thought. High NC individuals enjoy such thought and are apt to think deeply about events in their lives. In contrast, low NC individuals do not enjoy effortful thought and are relatively unmotivated to engage in this activity. Thus, the NC measure allowed us to identify participants who were likely to think deeply versus those who were likely to avoid such thought. Of course, it is *creative* thought that is associated with improved task performance [16, 24], and it should be noted that NC is not a measure of creativity per se. However, recent studies show that high (versus low) NC participants perform better on measures of creativity [6, 12]. It is likely that low NC individuals, being unlikely to engage in effortful thought, are at a disadvantage when it comes to discovering creative solutions to usability problems. If reasoning about potential solutions is a necessary condition under which attractive phones lead to improved task performance, we should observe that the attractiveness-performance relationship is stronger for high versus low NC participants.

In summary, the present research first and foremost tested whether attractiveness explained additional variance in usability ratings that was not explained by effectiveness and efficiency. We also conducted analyses testing whether our results were consistent with a popular framework proposed to explain the attractiveness-usability relationship [24]. We make no claims that our work validates the proposed mechanism or precludes the existence of other equally valid mechanisms. Rather, we examine whether Norman's framework can account for our data.

METHOD

Participants

Participants were 64 women and 42 men (age > 17) from the Kansas City metropolitan area. All participants owned a mobile phone and used it in their daily lives, but none had prior experience using the phone model they tested in the lab. Lab sessions lasted two hours. Each participant tested only one phone and received $75 upon completion.

Procedure

Participants attended sessions individually. Upon arrival, participants received the phone to be tested and a brief description of the experimental procedure. Next, they attempted to perform a specified set of common mobile phone tasks[1] using the assigned phone. Participants completed three blocks of these tasks. In each block, they completed the full set of tasks in random order. Participants rated the phone's usability twice – once after block 1 then again after block 3 (i.e., after extended practice) – and they completed scale measures of phone attractiveness and NC once following block 3.

All participants completed the same procedure under nearly identical conditions (i.e., same lab room, apparatus, experimenters). We note, however, that the phone model being tested varied across individuals. Ideally, we would have tested a single phone model to eliminate any variability in our results that might be attributed to differences between phones. However, it was necessary to

[1] All participants performed nine tasks: place a call, change the assigned ringer, add a contact, return last missed call, retrieve voicemail, retrieve text message, send text message, read a news article on the web browser, and take and send a picture. Depending on the phone being benchmarked, participants also performed one or two of the following tasks: download a ringer, play a song, make a push-to-talk call, and send email.

conduct this research in conjunction with business goals to collect data on a variety of phone models to establish a benchmark rating of usability for each device. In total, we used seven different phone models, each tested by a small number of participants (modal n = 12), as is common practice in usability testing in industry [21].

We aggregated across phone models to facilitate analyses. This is not uncommon in research investigating factors that shape mobile phone-related attitudes and behaviors. For instance, researchers have collapsed across phone models to identify factors that predict phone satisfaction [20] and to examine demographic and personality effects on mobile phone usage [7, 17]. Still, some readers might be concerned about whether it is appropriate to aggregate across phone models because relationships between perceived usability and attractiveness, effectiveness, and efficiency could differ across devices. To assuage these concerns we conducted several analyses to determine whether the relationships we examined in the present research differed as a function of important phone features. Specifically, we conducted regression analyses predicting usability ratings as a function of (a) a phone feature variable, (b) a predictor variable (i.e., effectiveness, efficiency, or attractiveness), and (c) the interaction between the phone feature and predictor. A statistically significant interaction would indicate that the relationship between usability ratings and the predictor variable differed for phones with (versus without) the phone feature being tested. No significant interactions emerged in any of these models testing the following phone features: (a) smartphone versus non-smartphone, (b) style of graphics shown on the main screen, or (c) iDen versus CDMA network technologies. Furthermore, in a separate study [27] we observed that the attractiveness-usability relationship was statistically significant for each of 33 phone models assessed. In this study, the attractiveness-usability relationship did not appear to vary systematically as a function of phone model. Given these results, we felt comfortable aggregating across phone models.

Measures

Effectiveness of Task Performance

For each phone task attempted during the session, the experimenter recorded whether the participant completed the task successfully (if the participant achieved the goal described in the task instructions) or unsuccessfully (if the participant did not achieve the task goal, exceeded a four-minute time limit to complete the task, or chose to stop trying before reaching the time limit). The participant received no feedback concerning success versus failure.

For each participant, we calculated the percent of tasks completed successfully during the session (i.e., the number of successful tasks divided by the total number of tasks). The percent of tasks completed successfully served as our measure of effectiveness.

Efficiency of Task Performance

For each phone task a participant completed successfully, the experimenter recorded the time required to complete the task (i.e., time on task). Participants received four minutes to complete each task, thus, times ranged from zero to 240 seconds. Times were not reported for unsuccessful tasks.

Time on task varied widely across tasks [28]. For instance, a time of 60 seconds was quite fast for the task of changing a ringer but quite slow for placing a call. To obtain an estimate of a participant's overall efficiency across many tasks, we converted time on task values to z-scores (i.e., deviations from the average time) within each task.[2] We then calculated the mean of a participant's z-scores to represent the extent to which that participant, on average, tended to complete tasks more quickly or more slowly than the typical participant. Negative values indicated faster-than-average performance and positive values indicated slower-than-average performance.

Perceived Usability of the Phone

Participants completed Brooke's [4] System Usability Scale (SUS) after blocks 1 and 3. They rated their agreement with the ten items of the SUS (e.g., "I thought this phone was easy to use") using a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We computed the mean rating across these ten items (after reverse-scoring negatively phrased items). The SUS achieved high reliability (Cronbach's α = .91 and .92, respectively, for blocks 1 and 3).

Phone Attractiveness

We developed a 7-item measure to assess participants' perceptions of a phone as attractive versus unattractive. With a set of four 7-point scales, participants rated the phone they used during the session as: (a) attractive versus unattractive, (b) beautiful versus ugly, (c) eye-catching versus plain, and (d) interesting versus boring. Participants also rated their agreement with three attractiveness-related statements (e.g., "I like the way this phone looks") using a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We computed the mean rating across these seven items such that higher values indicated greater attractiveness. This measure achieved high reliability (Cronbach's α = .90).

Need for Cognition

Participants completed the short version of the NC scale [9] by responding to 18 items (e.g., "I really enjoy a task that involves coming up with new solutions to problems") using a scale from 1 (*extremely uncharacteristic of you [not at all like you]*) to 5 (*extremely characteristic of you [very much like you]*). The scale achieved high reliability (Cronbach's α = .90).

---

[2] Time values for some tasks were positively skewed. For these tasks, we performed a natural log transformation to correct for non-normality prior to creating z-scores. Transformations were necessary for one task in block 1 and for eight tasks in block 3.

Variable	Mean	Standard Deviation	Range
Effectiveness, block 1	.70	.15	.27-1.00
Effectiveness, block 3	.83	.15	.40-1.00
Efficiency, block 1	.05	.54	-1.00-1.15
Efficiency, block 3	.04	.51	-1.14-1.19
SUS, block 1	4.59	1.32	1.30-7.00
SUS, block 3	5.01	1.28	1.30-7.00
Phone attractiveness	5.14	1.24	1.86-7.00
Need for cognition	3.62	.69	1.06-4.67

Table 1. Descriptive statistics.

RESULTS

Table 1 shows the variety of participants' performance-related experiences and perceptions of usability and phone attractiveness. Some participants performed well (e.g., 100% effectiveness) whereas others performed poorly (e.g., 27% effectiveness). Opinions differed widely regarding the attractiveness and usability of the devices tested (e.g., attractiveness ranged from 1.86 to 7.00). Thus, our analyses do not focus on a restricted range of experiences.

As can be seen in Table 2, high SUS ratings were associated with high effectiveness ($r_{\text{block 1}}$ = .51, $r_{\text{block 3}}$ = .33, *p*s < .001) and high efficiency ($r_{\text{block 1}}$ = -.33, $r_{\text{block 3}}$ = -.32, *p*s < .01). SUS ratings also were significantly higher for phones rated high (versus low) in attractiveness ($r_{\text{block 1}}$ = .53, $r_{\text{block 3}}$ = .50, *p*s < .001).

The Simultaneous Influence of Attractiveness, Effectiveness, and Efficiency on SUS

We constructed a hierarchical regression model estimating SUS as a function of two sets of variables tested in separate steps of the analysis. In the first step, we tested the impact of effectiveness and efficiency on SUS ratings. In the second step we added phone attractiveness to the model to determine whether inclusion of this variable improved our ability to predict SUS ratings above and beyond our ability to do so with effectiveness and efficiency alone. We omitted from the model several variables representing interactions among attractiveness, effectiveness, and efficiency because these interactions did not significantly predict SUS in preliminary analyses. Below, we report results of the hierarchical regressions separately for block 1 and block 3 of the session.

Block 1

In the first step of the hierarchical regression, block 1 effectiveness and efficiency together accounted for 29% of the variance in participants' block 1 SUS ratings (R^2 = .29). The addition of the attractiveness variable in the second step of the model resulted in a significant increase in R^2 to .51 ($F_{change}(1, 98) = 43.43$, $p < .001$). This means that attractiveness explained additional variance in usability ratings that was not explained by effectiveness and efficiency alone. In the hierarchical regression,

	SUS, block1	SUS, block3	Attractive-ness	Effectiveness, block1	Effectiveness, block3	Efficiency, block1	Efficiency, block3
SUS, block3	.86***						
Attractiveness	.53***	.50***					
Effectiveness, block1	.51***	.44***	.21*				
Effectiveness, block3	.27**	.33***	.22*	.66***			
Efficiency, block1	-.33***	-.23*	.08	-.36***	-.25**		
Efficiency, block3	-.34***	-.32**	-.11	-.47***	-.44***	.68***	
Need for cognition	.12	.14	.11	.13	.30**	.00	-.08

Note: *p<.05; **p<.01; ***p<.001.

Table 2. Correlations among variables.

attractiveness (b = 0.52, B = .48, $t(98)$ = 6.59, p < .001), effectiveness (b = 2.79, B = .32, $t(98)$ = 4.05, p < .001), and efficiency (b = -0.68, B = -.27, $t(98)$ = -3.54, p < .01) all were significantly associated with SUS ratings.

Block 3

In step 1 of the hierarchical regression, block 3 effectiveness and efficiency accounted for 14% of the variance in SUS ratings (R^2 = .14) and the addition of the attractiveness variable in the step 2 resulted in a significant increase in R^2 to .33 ($F_{change}(1, 98)$ = 28.39, p < .001). Again, including the attractiveness variable improved our ability to account for variability in SUS ratings. In this model, attractiveness (b = 0.47, B = .45, $t(98)$ = 5.33, p < .001) and efficiency (b = -0.55, B =-.22, $t(98)$ = -2.34, p < .05) were significantly associated with SUS ratings. The effectiveness-SUS relationship (which had been significant in step 1) was reduced to nonsignificance with the addition of the attractiveness variable.

The finding that the effectiveness-SUS relationship was reduced to nonsignificance when attractiveness was added to the regression model in block 3 suggests that effectiveness and attractiveness were statistically associated in our data. Below, we explore relationships between our three predictors in greater detail.

Relationships Between Attractiveness and Task Performance Measures

Attractiveness was associated with effectiveness such that participants achieved greater effectiveness using attractive versus unattractive phones ($r_{\text{block 1}}$ = .21, $r_{\text{block 3}}$ = .22, ps < .05; see Table 2). Attractiveness was not significantly associated with efficiency.

We conducted additional analyses to test whether our data were consistent with the theoretical framework proposed by Norman [24] in which improved task performance mediates the attractiveness-usability relationship. These analyses focused on effectiveness as a measure of task performance because, as stated above, attractiveness was unassociated with efficiency and, thus, efficiency cannot mediate the attractiveness-usability relationship [18].

To test whether effectiveness can explain why attractiveness was associated with SUS, we performed a set of analyses described by Baron and Kenny [3] to test for mediation.[3] According to these authors, "a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the criterion" (p. 1176). In the context of the present research, we tested the notion that effectiveness acts as a mediating variable that accounts for the relation between the predictor, phone attractiveness, and the criterion variable, perceived usability (SUS). Baron and Kenny specified four conditions that must be met for evidence of mediation: (1) the effect of the predictor (attractiveness) on the criterion (SUS) is statistically significant; (2) the effect of the predictor (attractiveness) on the mediator (effectiveness) is significant; (3) the effect of the mediator (effectiveness) on the criterion (SUS) is significant when controlling for the effect of the predictor (attractiveness); and (4) the effect of the predictor (attractiveness) on the criterion (SUS), when controlling for the effect of the mediator (effectiveness) on the criterion, is smaller than the direct effect of the predictor on the criterion assessed in condition 1.[4] In rare cases, testing condition 4 reveals that the effect of the predictor on the criterion is reduced to nonsignificance when controlling for the effect of the mediator. That is, the mediating variable completely explains why the predictor is associated with the criterion variable. This represents complete mediation. More common is the case in which a significant relationship still exists between predictor and criterion when controlling for the effect of the mediator (in condition 4), but this predictor-criterion relationship is reduced in magnitude compared to that observed in condition 1. This represents partial mediation and indicates that the mediating variable explains only part of the reason why the predictor is associated with the criterion variable.

Results described previously showed significant relationships consistent with conditions 1 and 2 of Baron and Kenny's model. To test conditions 3 and 4, we performed multiple regression analyses (for blocks 1 and 3, separately) predicting SUS as a function of both attractiveness and effectiveness.

Block 1

Using data from block 1, we observed a significant effect of effectiveness on SUS controlling for attractiveness (b = 3.72, B = 0.42, $t(99)$ = 5.53, p < .001). The effect of attractiveness on SUS, controlling for effectiveness, also was significant (b = 0.48, B = 0.44, $t(99)$ = 5.76, p < .001) and was smaller in magnitude than the direct effect of attractiveness on SUS without controlling for effectiveness (B = .53). Thus, the block 1 data were consistent with the hypothesis that effectiveness partially mediates the attractiveness-SUS relationship.

Block 3

Data from block 3 revealed a result similar to that described above. The effect of effectiveness on SUS controlling for attractiveness was significant (b = 2.17, B = 0.22, $t(99)$ = 2.57, p < .05). The effect of attractiveness on SUS

[3] We used this method because it is a straightforward procedure and is used most commonly in the psychological literature. Other methods exist to test for mediation and some authors advocate the use of the statistical technique of bootstrapping [31]. We obtained identical results using bootstrapping and Baron & Kenny's regression approach.

[4] We do not intend to imply that our correlational analyses allowed us to infer a causal path from attractiveness to effectiveness to perceived usability. Rather, we aimed to determine whether our data show relationships consistent with those described in Norman's theoretical framework.

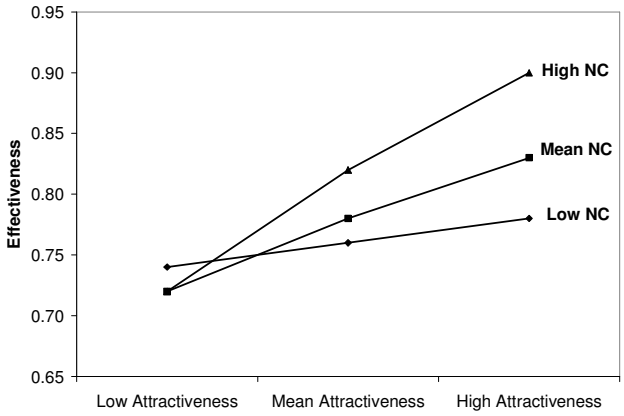

Figure 1. Effectiveness as a function of the Attractiveness X Need for Cognition (NC) interaction. Low and high values are 1 standard deviation away from the mean.

controlling for effectiveness also was significant (b = 0.47, B = 0.45, $t(99)$ = 5.22, p < .001) and was smaller than the direct effect of attractiveness on SUS (B = .50). As was the case in block 1, the block 3 data were consistent with the hypothesis that effectiveness partially mediates the attractiveness-SUS relationship.

The Role of Thought as a Moderator of the Attractiveness-Effectiveness Relationship

The analyses described above are consistent with the notion that high attractiveness leads to improved effectiveness, which in turn leads to high perceived usability. Next we conducted analyses to determine whether it is possible that effortful thought is the reason why attractiveness is associated with improved effectiveness. We conducted a multiple regression analysis predicting effectiveness as a function of attractiveness, NC, and the Attractiveness X NC interaction. Using block 1 effectiveness data we observed a significant interaction (b = 0.36, B = 1.45, $t(96)$ = 2.20, p < .05), which showed that the attractiveness-effectiveness relationship differed depending on participants' motivation to engage in effortful thought (see Figure 1). For low NC participants, who are unmotivated to engage in effortful thought, phone attractiveness had minimal impact on effectiveness. These participants were about equally effective regardless of whether they used phones they perceived to be attractive or unattractive. Indeed, among participants scoring lower than the median NC, attractiveness and effectiveness were not significantly related (r = .12, *ns*, n=49). In contrast, high NC participants, who are motivated to engage in effortful thought, performed better when using a phone they perceived as attractive versus unattractive. For participants who scored higher than the median NC, the attractiveness-effectiveness relationship was statistically significant (r = .28, p < .05, n = 51). This pattern is consistent with the notion that attractiveness leads to improved effectiveness only to the extent that attractiveness instigates thought about how to solve problems encountered during performance.

In a similar multiple regression using block 3 effectiveness data, the Attractiveness X NC interaction did not achieve statistical significance. However, the pattern of correlations for participants scoring higher than the median NC (r = .30, p < .05, n = 51) and those scoring lower than the median NC (r = .08, *ns*, n = 49) mirrored the findings for block 1 that showed a significant attractiveness-effectiveness relationship for high NC but not low NC participants.

DISCUSSION

When participants rated a phone's usability, their perceptions of the device appeared to be shaped by multiple factors. Consistent with past findings, the effectiveness and efficiency with which participants used the phone to perform common tasks significantly predicted their SUS ratings. The attractiveness of the phone's appearance had an independent effect on perceived usability, accounting for additional variance in SUS ratings above and beyond that which was explained by effectiveness and efficiency.

The relationship between attractiveness and SUS, controlling for the effects of effectiveness and efficiency, was not negligible. In fact, the magnitude of the attractiveness effect was comparable to (and even larger than) the effects of effectiveness and efficiency. Given this, it is plausible that high attractiveness could compensate for ineffective and inefficient performance, leading people to perceive a phone to be highly usable despite obvious usability problems.

Relationships with SUS Can Vary Over Time

Our results showed that the extent to which a given factor influences participants' perceptions of a phone's usability can vary over time. This was the case for effectiveness, which had a significant relationship with perceived usability (controlling for attractiveness and efficiency) in block 1 (B = .32) that was reduced to nonsignificance in block 3 (B = .13). Why might participants rely on information about their ability to effectively complete tasks as a basis for their perception of the phone's usability in block 1 but not block 3? One possible reason is that effectiveness was relatively low in block 1 (mean = 70%), thus, success versus failure at phone-related tasks was a highly salient indicator of the ease of using the phone. By block 3, practice rendered effectiveness much higher (mean = 83%). Indeed, 18% of participants effectively completed every task in block 3 (compared to only 3% in block 1). Plausibly, as participants became more adept at performing the tasks they began to expect success and might have taken for granted that they should be able to perform the tasks effectively. If so, effectiveness might have seemed less relevant to their evaluation of usability by block 3.

The strength with which efficiency and attractiveness influenced usability ratings remained stable across blocks 1 and 3 in the present research. However, other evidence

suggests these effects also vary under certain circumstances. Quinn and Weller [28] found that the influence of efficiency on perceived usability varied depending on effectiveness. The effect of efficiency was minimal at low levels of effectiveness and only shaped usability ratings when participants achieved average-to-high effectiveness. In addition, evidence from a longitudinal study [26] indicated that the attractiveness-SUS relationship became stronger after extended periods of phone ownership.

Why Is Attractiveness Associated With Usability?

Although the present research was not designed as a critical test of mechanisms underlying the attractiveness-usability relationship, we were able to examine relationships among several variables that make up one possible mechanism, proposed by Norman [24], which currently is popular among user experience professionals. Generally, our findings were consistent with Norman's framework. We found that participants achieved more effective task performance using attractive versus unattractive phones, which supports the notion that "attractive things work better" than unattractive things [24, p. 17]. Of course, there are multiple ways to measure how well a product works (e.g., error rate, quality of outcome, time on task) [1, 14] and it is unlikely that attractive products will outperform unattractive products on every measure of task performance [16]. Indeed, in the present research attractive products had no advantage over unattractive products in terms of efficiency.

Our results revealed an important role for cognition as a moderator of the attractiveness-effectiveness relationship, which is consistent with Norman's framework. We found that attractiveness was associated with effectiveness only among high NC participants (who are intrinsically motivated to engage in effortful thought) and not among low NC participants (who do not enjoy such thought and are relatively unmotivated to engage in it). Presumably, attractive products set the stage for effortful thought, but users must be motivated to engage in this kind of thought when the opportunity arises.

Finally, our data were consistent with the idea that effectiveness partially mediates the relationship between attractiveness and perceived usability, which fits with Norman's framework. Effectiveness did not completely account for the attractiveness-usability relationship, however, and a significant direct association existed between attractiveness and perceived usability. This suggests there are additional mechanisms underlying the relationship between these two variables.

Why did attractiveness have a direct (unmediated) effect on usability in the present research? Though the present data do not provide an answer, several possibilities are worth investigating in future research. Perhaps some other aspect of task performance that was unmeasured in our research (e.g., number of errors committed before effective completion) together with our effectiveness measure (percentage of tasks completed successfully) might completely mediate the path from attractiveness to perceived usability. Alternatively, other mechanisms which do not operate via actual improvements in task performance could explain the attractiveness-usability relationship. For example, Tractinsky, Katz, and Ikar [33] speculated that people hold expectations that attractive products, being well designed in the aesthetic sense, also will be well designed in terms of usability (i.e., a halo effect; see [23]).[5] Research and theory on *affect-as-information* provides another possible explanation. According to this perspective, affect can influence people's attitudes and evaluations in a variety of ways [11, 30]. During performance of a task, for example, subtle affective cues such as a user's positive mood "may be experienced as information about [her or his] own efficacy" [11, p. 461]. A person who is feeling especially capable and efficacious when asked whether a phone is usable might report that the phone is easier to operate than that same person would report at a time when her/his feelings of self-efficacy are not as strong. Another interesting possibility is that, in contrast to the notion that attractiveness causes usability, perhaps both attractiveness and usability are the result of a common characteristic of a phone's design. According to Reber, Schwarz, and Winkielman's [29] theory of beauty, objects that people perceive to be beautiful are those that possess characteristics that facilitate identification of the object (i.e., perceptual fluency). Characteristics such as symmetry, contrast and clarity, prototypicality, and repeated exposure make it easy for people to process information pertaining to an object and it is this ease of processing that creates the perception of an object as beautiful. Plausibly, design features that make a phone easy to process, identify, or recognize also confer task performance benefits such as efficiency or error reduction in addition to creating an attractive appearance.

In summary, our results were consistent with all of the components of Norman's proposed mechanism that we tested (note that we were unable to test the role of positive affect). However, our analyses leave open the possibility that other models could explain the attractiveness-usability relationship as well. Plausibly, alternative mechanisms operate in conjunction with those described by Norman to explain the attractiveness-usability relationship. Although future research is necessary to explicate the precise mechanisms linking attractiveness to usability, the present research adds to the growing literature showing the importance of product attractiveness for HCI and usability

[5] Although we raise this as a possible mechanism through which high attractiveness leads to high perceived usability without mediation by improved task performance, it also is possible that users' expectations that a product is easy to use instigate a self-fulfilling prophecy in which expectations trigger processes (e.g., greater confidence, persistence, or feelings of self-efficacy) that result in improved task performance.

testing. Clearly, attractiveness is one factor that must be considered when interpreting participant-rated usability.

REFERENCES

1. ANSI (2001). *Common industry format for usability test reports* (ANSI-NCITS 354-2001). Washington, DC: American National Standards Institute.
2. Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*, 114-123.
3. Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
4. Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189-194). London: Taylor & Francis.
5. Ben-Bassat, T., Meyer, J., & Tractinsky, N. (2006). Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transactions on Computer-Human Interaction, 13*, 210-234.
6. Butler, A. B., Scherer, L. L., & Reiter-Palmon, R. (2003). Effects of solution elicitation aids and need for cognition on the generation of solutions to ill-structured problems. *Creativity Research Journal, 15*, 235-244.
7. Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior, 24*, 346-360.
8. Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197-253.
9. Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.
10. Chawda, B., Craft, B., Cairns, P., Rüger, S., & Heech, D. (2005). Do attractive things work better? An exploration of search tool visualizations. *Proc. HCI 2005, 2*, 46-51.
11. Clore, G. L., & Schnall, S. (2005). The influence of affect on attitude. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 437-489). Mahwah, NJ: Lawrence Erlbaum Associates.
12. Dollinger, S. J. (2003). Need for uniqueness, need for cognition, and creativity. *Journal of Creative Behavior, 37*, 99-116.
13. Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? *CHI Letters, 2*, 345-352.
14. Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*, 79-102.
15. Hornbæk, K., & Law, E. L. (2007). Meta-analysis of correlations among usability measures. *Proc. CHI 2007,* 617-626.
16. Isen, A. M. (2001). An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications. *Journal of Consumer Psychology, 11*, 75-85.
17. Jackson, L. A., Zhao, Y., Kolenic, A., Fitzgerald, H. E., Harold, R., & Von Eye, A. (2008). Race, gender, and information technology use: The new digital divide. *CyberPsychology & Behavior, 11*, 437-442.
18. Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 233-265). New York: McGraw-Hill).
19. Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. *Proc. CHI 1995,* 292-293.
20. Ling, C., Hwang, W., & Salvendy, G. (2007). A survey of what customers want in a cell phone design. *Behaviour & Information Technology, 26*, 149-163.
21. Nielsen, J. (2000). *Why you only need to test with 5 users.* Retrieved December 14, 2009 from http://www.useit.com/alertbox/20000319.html
22. Nielsen, J., & Levy, J. (1994). Measuring usability: Preference versus performance. *Communications of the ACM, 37*, 66-76.
23. Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*, 250-256.
24. Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. New York: Basic Books.
25. Norman, D. A., & Ortony, A. (2006). Designers and users: Two perspectives on emotion and design. In S. Bagnara & G. Crampton Smith (Eds.), *Theories and practice in interaction design* (pp. 91-103). Mahwah, NJ: Erlbaum.
26. Quinn, J. M. (in preparation). *The aesthetic appeal of mobile phones: Users' ratings usually decline over time*.

27. Quinn, J. M. *Moderators of the attractiveness-usability relationship*. Unpublished data.

28. Quinn, J. M., & Weller, D. R. (2009). When users judge a mobile phone's usability, which features matter? *Proc. Usability Professionals' Association*.

29. Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*, 364-382.

30. Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.

31. Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.

32. Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. *CHI 97 Electronic publications: Papers.* http://sigchi.org/chi97/proceedings/paper/nt.htm

33. Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*, 127-145.