Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing

Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, Kristie Fisher

▶ **To cite this version:**

Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, Kristie Fisher. Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing. 14th International Conference on Human-Computer Interaction (INTERACT), Sep 2013, Cape Town, South Africa. pp.74-91, 10.1007/978-3-642-40477-1_5 . hal-01504930

HAL Id: hal-01504930

https://inria.hal.science/hal-01504930v1

Submitted on 10 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Nudging People Away From Privacy-Invasive Mobile Apps Through Visual Framing

Eun Kyoung Choe[1, 2], Jaeyeon Jung[1], Bongshin Lee[1], and Kristie Fisher[3]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
[2] The Information School, University of Washington, Seattle, WA 98195, USA
[3] Microsoft Studios, One Microsoft Way, Redmond, WA 98052, USA

eunky@uw.edu, {jjung, bongshin, kfisher}@microsoft.com

Abstract. Smartphone users visit application marketplaces (or app stores) to search and install applications. However, these app stores are not free from privacy-invasive apps, which collect personal information without sufficient disclosure or people's consent. To nudge people away from privacy-invasive apps, we created a visual representation of the mobile app's privacy rating. Inspired by "*Framing Effects,*" we designed semantically equivalent visuals that are framed in either a positive or negative way. We investigated the effect of the visual privacy rating, framing, and user rating on people's perception of an app (e.g., trustworthiness) through two experiments. In Study 1, participants were able to understand the intended meaning of the visual privacy ratings. In Study 2, we found a strong main effect for visual privacy rating on participants' perception of an app, and framing effects in a low privacy rating app. We discuss implications for designing visual privacy ratings, including the use of *positive visual framing* to nudge people away from privacy-invasive apps.
Keywords: Visual framing; privacy; privacy metrics; rating; nudge; framing effect; valence; positive framing; negative framing; Mechanical Turk.

1 Introduction

Application marketplaces (a.k.a. app stores) have become the mainstream channels to distribute applications onto smartphones. Over 700,000 apps have been published both in the Apple *App Store* and in *Google Play* as of Dec. 2012 [1,4]. However, this abundance of choice comes with consequences; app stores are not free from privacy-invasive apps that collect personal information without sufficient disclosure or user consent. An investigation of 100 popular apps (for iPhone and Android) shows that many of these apps collect personal information such as location and the phone identification numbers, and some apps share these data with third parties without proper user consent [28]. For example, Path, a popular social media app, was recently found to transmit the user's contacts stored on the phone without explicit permission [32].

When shopping for an app in the app store, people have an opportunity to compare different apps that provide similar functionality. For example, searching for "weather" returns more than 1,000 results in Google Play. Clicking an app of interest, people can view detailed information, such as description, screen shots, user ratings, and reviews of the app. In this detailed view, some app stores provide privacy-related information regarding which types of data the app may access; however, prior research has shown that people do not read the permission warnings, and even if they do, they do not understand what the information means, as the terms are vague and confusing [12,16]. Nonetheless, a recent survey reports that 30% of survey respondents had uninstalled apps found to collect personal information that they did not want to share [6].

The goal of this work was to explore novel ways to nudge people away from privacy-invasive apps when they search for and compare apps to install. Specifically, we created *visual representations* of an app's level of privacy protection and investigated how the visual representations influence people's perception of an app. Similar to a movie critics' rating [23], we created visuals for a *privacy critics' rating* of an app, conveying how privacy-preserving or privacy-invasive the app is (Figure 1).

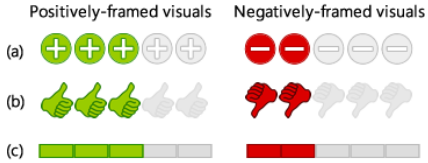

Fig. 1. Visual Framing. Positively-framed visuals of a rating of 3 (left column) is semantically equivalent to negatively-framed visuals of a rating of 2 (right column).

To create influential, persuasive visuals for privacy ratings, we leveraged the well-known "Framing Effects" [30]. The key idea is that people's decisions, in part, depend on the way problems are stated (e.g., positively or negatively). A classic example is how a doctor describes the odds of a grueling operation: many would prefer to choose an operation of which an outcome is "90 out of 100 are *alive* after five years" than "10 out of 100 are *dead* after five years" [21]. Even if these two phrases contain the same information, people—even experts (i.e., doctors)—are systematically subject to framing effects. In addition, framing effects occur without people knowing that they are being affected by it. People are susceptible to the framing as long as they understand the *valence* of an option—whether something is good or bad—without necessarily understanding what makes the option appealing.

Our goal was therefore *not* to make people *understand* the details of how an app's privacy rating is calculated or to help them make an informed decision. Rather, in this work, we investigated whether we can leverage framing effects to nudge people away from privacy invasive apps using visual representations. We used visual elements such as colors and symbols to make the valence information even more salient than text-only valence descriptions. To study the effect of the visual framing of a privacy rating, we first created semantically equivalent visuals for a privacy rating that highlight either the positivity or negativity of the rating, respectively. Then, we measured the effect of the positively- and negatively-framed visuals for privacy ratings on how

people perceive smartphone apps with/without a *user rating*. Privacy rating is just one aspect among many other attributes of an app whereas the user rating reflects how general audience holistically thinks about the app based on their individual experiences. Furthermore, people self-reported [8,11,12] that they consider privacy-related information far less important than user ratings in making app choices. Therefore, we considered the effect of the privacy rating in conjunction with the user rating.

Our contributions are threefold. First, we detail visual attributes (e.g., color) and semantics (e.g., valence, sign) that contribute to visual framing. Second, we investigate if framing effects transfer to visual representations of a privacy rating and shift people's perceptions of apps. Third, based on the lessons learned from two experiments, we discuss design implications for visualizing privacy ratings of smartphone apps in such a way to nudge people to avoid privacy-invasive apps.

2 Related Work

2.1 Mobile App Privacy and Permissions

Recent studies reveal the limitations of existing approaches to presenting privacy-related information on mobile phones. Looking into Android's permission interface, two studies show that people pay little attention to permission requests when installing an app and have a poor understanding of what each permission means [12, 16]. Semi-structured interviews of 20 Android users discover that participants do not understand Android permissions, which were described as "confusing, misleading, jargon-filled, and poorly grouped" [16]. An online survey of 308 Android users reports that only 17.5% of respondents reported looking at permissions during their last app installation [12]. Moreover, only 3% of respondents could provide correct answers for permission comprehension questions [12]. On the contrary, the same study reports that 71% of respondents looked at some type of user reviews before installing an app. Similarly, Chin and colleagues reveal that participants often rely on user ratings when deciding an app to install [8].

Realizing these limitations, researchers have made an effort to design simple and easy-to-understand visual representations of privacy-related information. For example, Kelley and colleagues show that presenting online privacy policies using standardized tables improves people' comprehension of the privacy policies [15]. Cranor and colleagues use a symbol (called Privacy Bird) to indicate whether a website's privacy policy matches a user's preference [9]. A follow-up study designs "privacy icons" and demonstrates that when the privacy icons were presented, participants were willing to pay a premium to purchase items from the websites that better protect users' privacy [29]. Similar to these studies, we created simple visual representations for privacy information. However, our work is different from the previous work as our goal was not to improve people's comprehension of the exact details of privacy policies, but rather to create visual representations that can affect how people perceive an app and as a result can nudge them away from privacy-invasive apps.

One study by Lin and colleagues [20] addresses the limitations of Android's permission interface using crowdsourcing. The authors propose a privacy summary that highlights the use of permissions that did not match other people's expectations. To capture people's expectations, the authors conducted online surveys, recruiting respondents from Amazon Mechanical Turk. Such measured expectations can be useful for creating an app's privacy rating, which our work could leverage. In comparison, our work examines *framing effects* in *visual representations* for privacy information with the aim to identify what makes certain visuals more influential and persuasive than others and leverage those properties to influence user perceptions.

2.2 Theoretical Framework: Framing Effects

A few recent studies apply behavioral economics theories in designing persuasive technologies. Lee and colleagues [17] apply three persuasion techniques drawn from behavioral economics—the default option strategy, planning strategy, and asymmetric choice strategy—in designing choices for healthy eating. Inspired by their work, we explore a novel approach to employ *framing effects* to nudge people away from privacy-invasive apps. Humans have been thought to make rational choices, of which the core principles include *invariance* [31]. Invariance requires that two versions of a choice that are recognized to be equivalent should yield the same preference regardless of the manner in which they are described. However, Tversky & Kahneman [30] reveal that presenting the same option can alter people's decisions when varying the framing of acts, contingencies, or outcomes.

Since Tversky & Kahneman first discussed framing effects explaining how valence influences people's willingness to take risks, framings have been studied using *text descriptions* in many domains including privacy domain [13]. To better understand when and why different types of framing will have an effect, Levin and colleagues develop a typology of framings, and distinguish between three different kinds of framings—risky choice, attribute, and goal framing [19]. Our work is particularly inspired by the *attribute framing*, which manipulates the valence of a *single* attribute within a given context. For example, an attribute of ground beef can be labeled as either "75% lean" or "25% fat," and a study shows that people favor the former even though the two labels convey the same information [18]. Framing effects occur because people tend to be somewhat mindless when consuming information in daily life [27]. Therefore, when applied properly, framing can be used as a powerful nudge to influence people to make a better decision [27].

2.3 Nudging for Privacy Decision

Acquisti & Grossklags discuss the relevance of framing to privacy research [2]. In [3], Acquisti coins the term, "nudging privacy," and discusses how existing research in behavioral economics and psychology can help better understand privacy decision making. Furthermore, Balebako and colleagues broadly discuss how their ongoing work in location sharing can leverage nudging interventions to help people make better privacy decisions [5].

In HCI research, the term "nudging" has been used in the broad sense outside of behavioral economics as a means to influence people's behaviors, such as shopping [14] or exercise habits [25]. When designing a system whose goal is to nudge people toward a certain direction, the system might undermine the decision making power of individuals. Thaler & Sunstein [27] state that a nudge should be "easy and cheap to avoid." Thus, banning privacy-invasive apps entirely from app stores is not a nudge. We also note that the use of "framing" as a design approach for nudging might impose ethical concerns as discussed in [7]. We share the same vision with the prior research mentioned above in that leveraging framing can be a promising approach to nudge people toward better privacy decisions. However, these techniques should be applied with care so as not to limit people's freedom to choose.

3 Research Questions and Experimental Setups

We explore the following research questions (RQ) through two experiments:

> RQ1. Can we create complementary visual framings that convey semantically equivalent privacy rating information?
>
> RQ2. Do complementary privacy ratings have similar influence on how people perceive an app? If not, is a negative visual framing more effective in nudging people away from privacy-invasive apps?
>
> RQ3. Do people's perceptions of an app change if a privacy rating is accompanied by a user rating?

For each study, we created an online experimental setup and recruited participants using Amazon Mechanical Turk (MTurk). We made sure no one could participate more than once by setting a cookie in participants' browser and by removing answers from duplicate MTurk IDs. The studies were available only to U.S. and Canada residents with at least a 95% approval rate (through a screening option that MTurk provides). We compensated MTurk participants $0.50 USD per survey for Study 1, and $1 USD per survey for Study 2.

4 Study 1: Creating Complementary Visuals

As a first step toward identifying visuals that have higher nudging power, we investigated whether it is viable to create complementary visual framings that convey semantically equivalent privacy rating information. We designed two sets of icons: positively-framed (PF) icons using a green plus sign (+) and negatively-framed (NF) icons using a red minus sign (-) (Figure 1-a). Because most of the ratings (e.g., Amazon star rating) use PF icons (i.e., the more stars, the better), we speculated that people were already familiar with the PF icons but not with the NF icons. We measured the level of comprehension of the PF and NF icons after short training. Similar to prior framing studies, we conducted a between-subjects experiment with two groups: PF icon group and NF icon group.

4.1 Survey Content

We created online surveys for PF and NF conditions. The surveys consisted of a privacy rating description page and two sets of eight icon comparison questions (the first set was for training). On the first page, we introduced a hypothetical privacy metric called "Privacy Critics' Rating." To help participants understand valence of the metric, we provided the following explanation:

> Imagine you are deciding whether to install an app on your smartphone. **Privacy experts** ran a series of tests to assess the app's level of privacy. These tests measured how likely the app is to capture your personal information and to share/sell your information to advertisers or other partners.

We manipulated the description and a screenshot of a fake game app, "Tic Tac Toe," in the following manner:

PF condition: We explained that the privacy rating scale represents the proportion of tests that the app **passed**. Therefore, if an app has more green-colored (PF) icons in the rating, it is more privacy-preserving. The screenshot included a privacy rating of 4 (out of 5).

NF condition: We explained that the privacy rating scale represents the proportion of tests that the app **failed**. Therefore, if an app has more red-colored (NF) icons in the rating, it is more privacy-invasive. The screenshot included a privacy rating of 1 (out of 5).

According to our description, a privacy rating of 4 (out of 5) in the PF condition is equivalent to a privacy rating of 1 (out of 5) in the NF condition. After the description, we asked participants to answer two sets of eight icon comparison questions *as accurately and quickly as possible*. We showed two different privacy ratings and asked which of the two privacy ratings is more privacy-invasive or privacy-preserving (Figure 2). These two questions appeared in a random order. We placed one question per page; participants had to click a "next" button in order to proceed to the next question, which allowed us to measure a task completion time for each question.

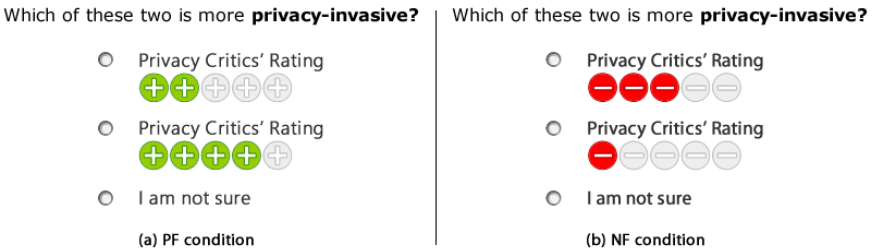

Fig. 2. An example of the icon comparison questions and the complementary visuals for the PF (a) and NF (b) conditions.

4.2 Measures

The dependent measures were accuracy (i.e., the number of correct responses) and task completion time (i.e., average response time per question). The first set of eight questions was a training set, thereby using only the second set of eight questions for the analysis. We also collected qualitative feedback in free-form text.

4.3 Participants

We recruited 129 participants and randomly assigned them to either the PF condition ($N = 67$) or NF condition ($N = 62$). We limited the study participants to mouse users because different input methods could influence task completion time significantly. Additionally, we filtered MTurkers who did not pay careful attention to the study instructions as indicated by the description reading time; we removed 47 participants who spent less than 20 seconds on the description page (it took the authors more than 20 seconds simply to read the description).

4.4 Results of Study 1

The independent samples t-test showed that in terms of accuracy, there was not enough evidence to suggest that PF icon group ($N = 43$, $M = 7.81$, $SD = .59$) differs from NF icon group ($N = 39$, $M = 7.59$, $SD = 1.29$), $t(80) = 1.03$, $p = .31$. To compare the task completion time (in second), we first removed outliers defined by the mean minus two standard deviations ($M - 2SD$) (i.e., the number of correct responses < 5.7). Then we excluded incorrect responses in calculating the task completion time. The independent samples t-test showed that in terms of task completion time, there was not enough evidence to suggest that PF icon group ($N = 39$, $M = 3.29$, $SD = .92$) differs from NF icon group ($N = 36$, $M = 3.50$, $SD = 1.12$), $t(73) = -.89$, $p = .37$.

4.5 Discussion of Study 1

After reading the description and solving eight icon comparison questions, the majority of participants in both groups were able to comprehend privacy ratings in a similar manner. The PF and NF icons resulted in the comparable level of comprehension and speed by survey participants. Thus, we chose to use this set of PF and NF icons (Figure 1-a) as stimuli for our subsequent framing study.

We note, however, that even if there was no significant difference in the performance, 7 out of 36 (= 19%) participants in the NF icon group mentioned in their written comments that the NF icons were confusing. We suspect that this is due to a strong preconception of "*the more, the better,*" which is what the prevalent PF rating scale conveys. At the same time, other participants commented that the red minus sign we used to convey the negativity was particularly helpful to mark privacy-invasive apps. In Study 2, we explored whether the complementary visuals have similar influence on how people perceive an app.

5 Study 2: Positive vs. Negative Framings

Given that people can comprehend the positive and negative visual framings in a similar manner with some training, we investigated whether there is any effect of *framing* on how people perceive apps. In prior framing studies using text descriptions, researchers report that negatively-framed information tends to influence *more* than the positively-framed information of the same magnitude [24,26]. In Study 2, we tested which visual framing—positive or negative—was more effective in nudging people away from privacy-invasive apps using the icons we designed in Study 1. We also explored how people's perception of an app changes if a privacy rating of the app accompanies a user's overall rating (user rating). This resulted in a 2 (framing: PF icon; NF icon) x 3 (privacy rating: high; medium; low) x 2 (user rating: with a user rating of 3; without a user rating) mixed design with repeated measures; *framing* and *user rating* were between-subjects factors and *privacy rating* was a within-subjects factor, thereby forming four conditions: PF with & without a user rating of 3, and NF with & without a user rating of 3.

5.1 Survey Content

We created online surveys for the four conditions. The surveys consisted of three sections: (1) evaluating four apps (one dummy app followed by three apps of varying degrees of privacy ratings), (2) eight Privacy Critics' Rating icon comparison questions, and (3) demographic questions. At the beginning of the surveys, we introduced a hypothetical privacy metric called "Privacy Critics' Rating" using the same blurb from Study 1. Then, we showed a fake weather app with the following description and a screenshot.

> Imagine that you need a weather app, so you are searching the marketplace on your phone. While looking around many weather apps, you come across the following app page:

The screenshot contained a visual privacy rating, icon, and app description (Figure 3). We manipulated the screenshot of the weather app in the following manner:

Privacy Rating: For each participant, we showed a series of four screenshots of a weather app: one dummy set (user rating: 3.5; privacy rating: 2) and three levels of privacy ratings (high; medium; low) in a random order (Figure 4). The dummy set was included at the beginning to account for a longer reading time and familiarity with the screenshot and the question types.

Positive Framing (PF) vs. Negative Framing (NF): We created a PF and NF condition by manipulating the description of the Privacy Critics' Rating and its icon design as in Study 1 (Figure 4).

With a User Rating of 3 (w/ UR) vs. Without a User Rating (w/o UR): We chose a user rating of three as a representative case for the w/ UR condition because most of the popular free apps in the marketplace have an equal or higher user rating than three. Three green-colored stars (out of five) appeared right above the privacy rating

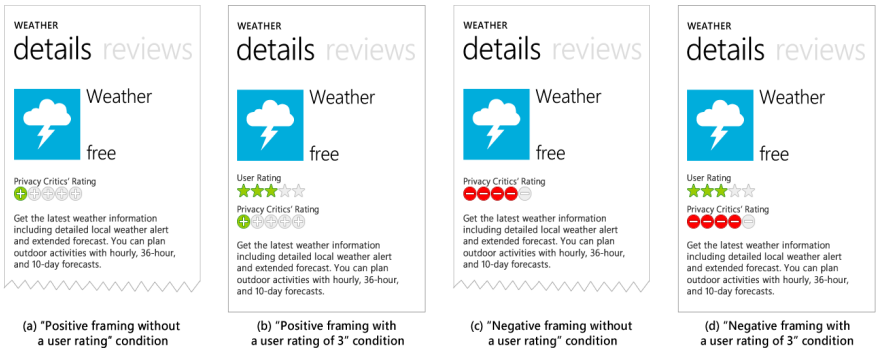

Fig. 3. Screenshots of the detailed view. Each screenshot represents the following: (a)—PF & w/o UR, (b)—PF & w/ UR, (c)—NF & w/o UR, and (d)—NF & w/ UR. Below the privacy rating, we provided descriptions of an app.

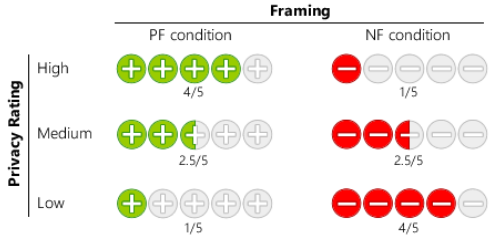

Fig. 4. This illustrates how we manipulated the privacy rating (3 levels) and framing (2 levels).

in the w/ UR conditions (Figures 3-b/d). No user rating was provided in the w/o UR conditions (Figures 3-a/c).

5.2 Measures

After showing each app, we measured people's perception toward each app by asking the following four questions: (1) trustworthiness of the app (TRUST), (2) likeability of the app (LIKE), (3) willingness to install the app (INST), and (4) willingness to recommend the app to a friend (RCMD). We measured TRUST because we suspected that people would associate an app's privacy rating with its trustworthiness. LIKE is a conventional measure in prior framing studies where a product's attribute is framed differently [19]. We measured INST and RCMD to gauge people's willingness to adopt an app. TRUST and LIKE were measured on a 7-point Likert scale, where 1 = not at all trustworthy / I strongly dislike this app, and 7 = very trustworthy / I strongly like this app. INST and RCMD were measured using Yes/No dichotomous questions. In addition, we suspected that participants' interest level toward the weather app could be a factor related to other dependent measures. So we measured self-reported interest level (INTEREST) toward the weather app on a 7-point Likert scale, where 1

= not at all interested, and 7 = very interested, at the very beginning of the survey as we showed the screenshot *without* the privacy and user rating information.

Within each condition, the privacy rating was the only factor that we varied. It was necessary for participants to understand how the privacy rating worked so that they could answer the evaluation questions based on correct understanding of the stimuli. Therefore, we measured accuracy and task completion time for the eight privacy rating icon comparison questions as in Study 1. However, this time, we placed these questions *after* the app evaluation questions to use them for *filtering* purpose rather than *training* purpose. We made this decision based on our pilot study and previous research indicating that framing effects are susceptible to rational thinking (e.g., asking people to provide a rationale for their choice eliminates or reduces the framing effects [22]).

5.3 Participants

We recruited 332 participants from Amazon Mechanical Turk and randomly assigned them to one of the four conditions: PF & w/o UR ($N = 75$); PF & w/ UR ($N = 95$); NF & w/o UR ($N = 79$); and NF & w/ UR ($N = 83$). We removed 18 duplicate participants (i.e., who participated in Study 1 or pilot studies). Using the same criteria from Study 1, we removed 79 participants who spent little time reading the descriptions. Among the remaining 235 participants, 55% of the participants ($N = 129$) were male, and 89% of the participants ($N = 210$) claimed that they own a smartphone.

5.4 Results of Study 2

After removing 21 participants from the analysis whose number of correct responses to the eight filtering questions was less than $M - 2SD$ (i.e., the number of correct responses < 5.3), we observed that participants' initial interest level toward the weather app was highly related to how much they trust the weather app, $F(1, 209) = 9.02$, $p = .003$, and how much they like the weather app, $F(1, 209) = 23.20$, $p < .001$. Therefore, TRUST and LIKE were analyzed using a mixed-design analysis of covariance (ANCOVA) controlling for the INTEREST as the covariate (Table 1). INST and RCMD were analyzed using a Pearson chi-square test (Table 2).

Table 1. This table shows strong main effects of privacy rating on TRUST and LIKE, and the framing effect (significant interaction effect between privacy rating and framing) on TRUST.

	TRUST (F-value)	LIKE (F-value)
Privacy Rating	55.55**	30.42**
Framing	2.33	.91
User Rating	1.83	3.41‡
Privacy Rating X Framing	7.48*	.83
Privacy Rating X User Rating	3.19‡	12.02**

**p<.001, *p<.05, ‡p<.10

"TRUST" Question. After controlling for the effect of INTEREST, we found a significant main effect of privacy rating on TRUST, $F(1.38, 289.13) = 55.55$, $p < .001$. Planned contrasts revealed that a high privacy rating app ($M = 5.15$) was regarded as more trustworthy than a medium privacy rating app ($M = 3.03$), $F(1, 209) = 50.39$, $p < .001$. Also, a medium privacy rating app was regarded as more trustworthy than a low privacy rating app ($M = 1.75$), $F(1, 209) = 29.60$, $p < .001$.

We found a significant interaction effect between privacy rating and framing, $F(1.38, 289.13) = 7.48$, $p = .003$ (Figure 5). To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across PF and NF conditions. We found a significant interaction between a high privacy rating app and a low privacy rating app across PF and NF conditions, $F(1, 209) = 9.09$, $p = .003$, and between a medium privacy rating app and a low privacy rating app across PF and NF conditions, $F(1, 209) = 8.89$, $p = .003$. Also, there was a marginally significant interaction between a high privacy rating app and a medium privacy rating app across PF and NF conditions, $F(1, 209) = 3.50$, $p = .06$. As Figure 5 indicates, the difference between PF ($M = 1.44$) and NF ($M = 2.05$) was significant for a low privacy rating app, $t(211) = 2.42$, $p = .02$. No significant differences were found between PF ($M = 5.27$) and NF ($M = 5.02$) conditions in a high privacy rating app, $t(211) = .34$, $p = .12$, or between PF ($M = 2.97$) and NF ($M = 3.10$) conditions in a medium privacy rating app, $t(211) = 1.58$, $p = .74$.

We found a marginally significant interaction between privacy rating and presence of user rating, $F(1.38, 289.13) = 3.19$, $p = .06$. This indicates that user rating might have different effects on app's trustworthiness at different levels of privacy rating. To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across w/ and w/o UR. We found a significant interaction when comparing a high privacy rating app to a medium privacy rating app across w/ and w/o UR, $F(1, 209) = 4.43$, $p = .04$. Also, we found a marginally significant interaction when comparing a high privacy rating app to a low privacy rating app across w/ and w/o UR, $F(1, 209) = 3.44$, $p = .07$. As Figure 6 indicates, when a user rating of 3 is shown, there is a decline in TRUST for a high privacy rating app while there is an increase in TRUST for a low and medium privacy rating app.

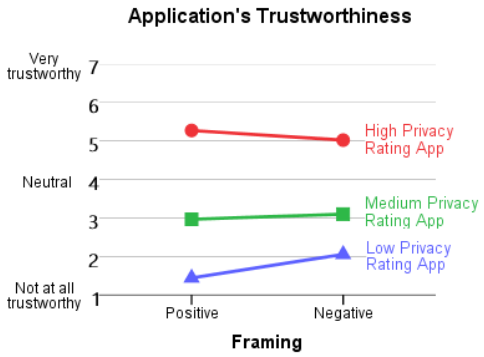

Fig. 5. This shows a significant interaction between privacy rating and framing on TRUST. Participants interpreted the NF icons more positively than the PF icons of the equivalent rating.

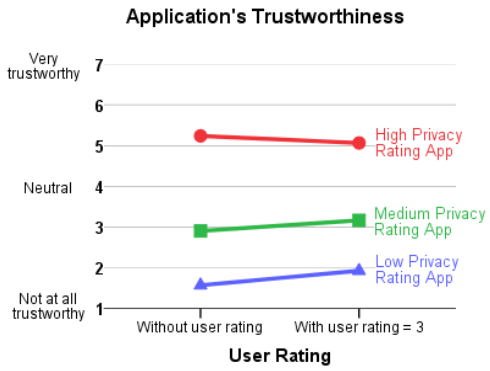

Fig. 6. This shows a marginally significant interaction (p = .06) between privacy rating and user rating on TRUST.

"LIKE" Question. After controlling for the effect of INTEREST, we found a significant main effect for privacy rating on LIKE, $F(1.46, 305.42) = 30.42$, $p < .001$. This effect tells us that how much participants liked the weather app was different for high, medium, and low privacy rating apps. Planned contrasts revealed that participants liked the high privacy rating app (M = 5.01) significantly more than the medium privacy rating app (M = 3.45), $F(1, 209) = 12.41$, p = .001. Also, participants liked the medium privacy rating app significantly more than the low privacy rating app (M = 2.17), $F(1, 209) = 48.01$, $p < .001$.

We found a marginally significant main effect for user rating, $F(1, 209) = 3.41$, p = .07. Participants liked the app w/ UR of 3 (M = 3.66) more than the app w/o UR (M = 3.42).

We found a significant interaction between privacy rating and user rating, $F(1.46, 305.42) = 12.02$, $p < .001$ (Figure 7). This indicates that user rating had different effects on LIKE depending on different levels of privacy rating. To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across w/ UR and w/o UR. We found a significant interaction when comparing a high privacy rating app to a medium privacy rating app across w/ UR and w/o UR, $F(1, 209) = 15.45$, $p < .001$. Also, we found a significant interaction when comparing a high privacy rating app to a low privacy rating app across w/ UR and w/o UR, $F(1, 209) = 14.02$, $p < .001$. The interaction graph (Figure 7) tells us that when a user rating of 3 is shown, there is a slight decline in LIKE for a high privacy rating app while there is an increase in LIKE for a low and medium privacy rating app. The remaining contrasts revealed no significant interaction when comparing a medium privacy rating app to a low privacy rating app in the w/o UR and w/ UR conditions, $F(1, 209) = 2.59$, p = .11.

"INSTALL" Question. We found a marginally significant association between framing and participants' choice of installing a low privacy rating app, χ^2 (1, N = 214) =

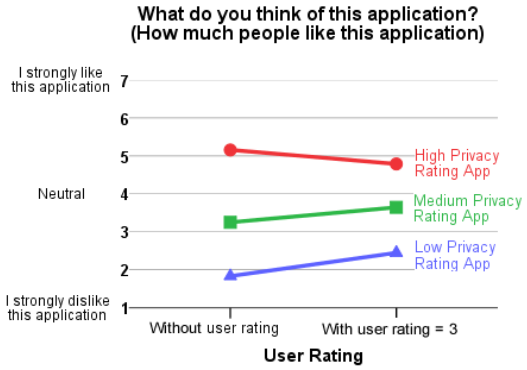

Fig. 7. This shows a significant interaction between privacy rating and user rating on LIKE. A user rating of 3 had a different effect on the high privacy rating app in comparison to medium/low privacy rating apps.

3.42, $p = .06$ (Table 2). The odds ratio implies that the odds of participants installing a low privacy rating app were 3.36 times higher if the rating were negatively framed than positively framed.

We found a significant association between framing and participants' choice of installing a medium privacy rating app, χ^2 (1, $N = 214$) = 5.17, $p = .02$ (Table 2). The odds ratio implies that the odds of participants installing a medium privacy rating app were 2.21 times higher if the rating were negatively framed than positively framed.

We found a marginally significant association between a user rating and participants' choice to install a high privacy rating app, χ^2 (1, $N = 214$) = 3.20, $p = .07$ (Table 2). The odds ratio implies that the odds of participants installing a high privacy rating app were 1.80 times higher if there were no user rating than the app accompanying a user rating of 3.

Table 2. More people answered that they would install / recommend a negatively-framed low privacy rating app than the same app that is positively-framed. The number of people who answered that they would install / recommend a high privacy rating app decreased when a user rating of 3 was shown than when no user rating was shown.

Privacy Rating	Answer	INSTALL Question			RECOMMEND Question		
		Positive Framing	Negative Framing	p-value[a]	Positive Framing	Negative Framing	p-value[a]
Low Privacy Rating App	Yes	3 (2.8%)	9 (8.6%)	.06‡	2 (1.8%)	10 (9.5%)	.02*
	No	106 (97.2%)	96 (91.4%)		107 (98.2%)	95 (90.5%)	
Medium Privacy Rating App	Yes	13 (11.9%)	25 (23.8%)	.02*	10 (9.2%)	17 (16.2%)	.12
	No	96 (88.1%)	80 (76.2%)		99 (90.8%)	88 (83.8%)	
		No User Rating	User Rating = 3	p-value[a]	No User Rating	User Rating = 3	p-value[a]
High Privacy Rating App	Yes	88 (82.2%)	77 (72.0%)	.07‡	83 (77.6%)	69 (64.5%)	.04*
	No	19 (17.8%)	13 (28.0%)		24 (22.4%)	38 (35.5%)	

a. A p-value is calculated from a Pearson's Chi-square test. *p<.05, ‡p<.10

"RECOMMEND" Question. We found a significant association between framing and participants' choice of recommending an app with low privacy rating, χ^2 (1, N = 214) = 5.97, p = .02 (Table 2). The odds ratio implies that the odds of participants recommending a medium privacy rating app to a friend were 5.53 times higher if the ratings were negatively framed than positively framed.

We also found a significant association between user rating and participants' choice of recommending an app with high privacy rating, χ^2 (1, N = 214) = 4.45, p = .04. The odds ratio implies that the odds of participants recommending a high privacy rating app were 1.90 times higher if there were no user rating than a user rating of 3.

5.5 Discussion of Study 2

Study 2 results show a strong effect of privacy rating on all dependent measures. This indicates that when a privacy rating of a given app is disclosed *visually*, people are influenced by the privacy rating. The influence of the privacy rating appears to decline (although still significant) when we showed a user rating of 3 (Figures 6/7). This suggests that people are susceptible to both privacy rating and user rating.

The effect of framing was subtle. First, we observed framing effects on TRUST in a low privacy rating app. Participants expressed a lower level of trustworthiness of an app when its privacy rating was positively framed than negatively framed. For medium and high privacy rating apps, framing effect did not occur. A similar trend was observed for INST and RCMD—a low privacy rating app was a common denominator for the framing effects to be observed, and when observed, it was always the negatively framed icons that people interpreted more positively. However, there was no framing effect on LIKE; after controlling for people's app interest level, privacy rating and user rating dominantly influenced LIKE. As we suspected, it appears that participants associated the privacy ratings with TRUST more than LIKE.

Prior framing studies using text descriptions consistently show that positive framing leads to more favorable evaluations than negative framing [e.g., 18,21]. Researchers demonstrate that describing an option in a negative light (e.g., "mortality rate") focuses attention on the unfavorable possibilities associated with this option, rendering it less acceptable to the decision-maker [19]. Therefore, we initially suspected that emphasizing negativity (e.g., privacy-invasiveness) would nudge people away from privacy invasive apps with a low privacy rating. However, our study results suggest this is not the case.

On the contrary, PF icons were more effective in making a low privacy rating app look more unfavorable. We suspect that people have strong connotations of "the more, the better" in the rating context. Because a negatively framed privacy invasive app has *more* signs in the rating than the equivalent PF icons (i.e., four minus signs vs. one plus sign), it is plausible that the higher number of ratings, regardless of its meaning, could have contributed to how people perceive the PF/NF icons.

Our results also suggest that there was no framing effect in the high and medium privacy rating apps. Therefore, the use of PF icons for depicting privacy ratings is a better choice for nudging people away from privacy invasive apps while not affecting high and medium privacy rating apps.

6 Discussion and Future Work

Our results suggest that visual representations of privacy information of apps can influence installation decisions by smartphone users. When disclosed visually on the detailed view of an app, the majority of participants commented that they found the privacy rating very helpful in deciding whether to install an app. In this section, we discuss design recommendations for a visual privacy rating as we answer our research questions. We also discuss limitations of this work and areas for future research.

6.1 Design Recommendations

To design complementary visuals that represent the same privacy rating in a positive or negative way (RQ1), we suggest leveraging visual attributes (e.g., colors) and semantics (e.g., valence, signs, symbols) that are prevalent throughout the culture for conveying valence. In text, people make favorable and unfavorable associations with positively and negatively phrased attribute labels (e.g., "fat" is associated with "bad" and "lean" with "good"). Similarly in visuals, we can take advantage of people's preconceptions built on their life experiences.

In designing PF and NF privacy ratings, we tried out various symbolic figures in red and green and chose the most promising one based on pilot tests measuring task completion time and error rate. For example, plus and minus signs, or thumbs-up and thumbs-down are well-known symbols in the US (Figures 1-a/b). When colors (red for negative and green for positive) are added on top of these, the valence becomes more pronounced. However, we ruled out thumbs-up and thumbs-down icons because they are already being used to convey different meanings elsewhere (e.g., "thumbs-up" is associated with "like" in Facebook). We also ruled out neutral figures (e.g., rectangles in Figure 1-c)—providing extra information (e.g., a legend) shall be necessary to frame neutral figures.

Study 2 results show that the complementary visuals do not have the same influence on how people perceive an app in certain conditions (RQ2). If the goal is to nudge people away from privacy-invasive apps, we suggest using a *positive visual framing* for privacy rating. However, this might not be the case outside the realm of "rating scale" design; when "the more, the better" connotation no longer exists, positive visual framing might lead to more favorable evaluations and vice versa, similar to prior framing effects shown in texts. Therefore, future research is warranted to study the influence of positive and negative visual framings in other contexts.

The results of Study 2 also show that people's perception of an app changes if a privacy rating is accompanied by a user rating but that a privacy rating still has a strong main effect (RQ3). This is promising in that people do understand and consider the level of privacy protection when it is presented to them visually. Prior research shows that textual privacy-related information is easily discarded [12,16]. This again suggests the need for *visual privacy ratings*. This work, however, does not address whether the main effect of visual privacy rating is due to the privacy information shown in a rating scale, due to the visuals, or due to the combination of both. Therefore, future research is needed to tease out these factors.

6.2 Limitation and Future Work

In running online experiments with participants recruited from the MTurk site, we made a conservative assumption of filtering mindless Turkers defined by those who spent less than 20 seconds on the description page where we explained the privacy rating and what it represents. Because we introduced a privacy rating with which most people are unfamiliar, eliminating answers from an arbitrary guess was necessary. This is analogous to avoiding recruits who do not understand what "fat" and "lean" mean in the ground beef framing study [18]. By the same token, we eliminated outliers whose number of correct responses to the icon comparison questions was less than $M - 2SD$. However, this does not guarantee that those who spent longer than 20 seconds on the description page correctly interpreted the privacy rating, which is beyond our control when running an experimental study online.

In this work, it was not our main interest to investigate the relative influence of privacy-related information in comparison to user ratings (although we touched upon the interaction between a privacy rating and user rating in Study 2). However, our study results suggest that understanding this topic would inform how much emphasis should be placed visually on the privacy ratings in comparison to user ratings. To answer this question, we would need to understand how people's perceptions of an app change depending on different designs and the relative weight of a visual privacy rating in return.

We would also like to further investigate how visual framings can be applied outside the privacy domain as a method for nudging. Health information feedback of consumer health devices, for example, is designed with a specific goal in mind—promoting health-enhancing behaviors. One promising direction is to design visual framings that represent the outcome of a healthy or unhealthy behavior stressing either the positive or negative consequences as we provide health information feedback. Framing research in texts has addressed when framing effects are *reduced* or *eliminated* [10,22], which helps HCI researchers select application areas for employing the visual framing.

7 Conclusion

We created visuals for a mobile app's privacy rating by leveraging the well-known "Framing Effects." In Study 1, we showed that it is viable to create semantically equivalent privacy ratings framed in either a positive or negative light. In Study 2, we showed a strong main effect for visual privacy rating on people's perception of an app. Our study results suggest that when an app's privacy property is provided in the form of a visual privacy rating, people are able to understand and are heavily influenced by it. We also examined how the visual framings shift people's perception of an app, and observed the occurrence of framing effects in a low privacy rating app. In designing a visual privacy rating to nudge people away from privacy-invasive apps, we recommend using positive visual framing, and leveraging visual attributes and semantics that are prevalent throughout the culture for conveying valence. Investigating the relative influence of a visual privacy rating in comparison to a user rating

warrants future research efforts. In closing, this work provides empirical guidance for creating influential, persuasive visual framing that can nudge people away from privacy invasive apps.

Acknowledgment

We would like to thank Tim Paek (Microsoft Research) for his thoughtful comments on earlier drafts of this paper. We also thank Stuart Schechter (Microsoft Research) for providing suggestions on the study design.

References

1. 148Apps.biz. (2012). App Store Metrics. Retrieved Dec. 31, 2012, from http://148apps.biz/app-store-metrics.
2. Acquisti, A., & Grossklags, J. (2007). What can behavioral economics teach us about privacy? In A. Acquisti, S. Gritzalis, S. Di Vimercati, C. Lambrinoudakis (Eds.), *Digital Privacy: Theory, Technologies, and Practices* (pp. 363–379). Auerbach Publications.
3. Acquisti, A. (2009). Nudging privacy: behavioral economics of personal information. *Security & Privacy, 7*(6), 82–85.
4. AndroLib. (2012). Retrieved Dec. 31, 2012, from http://www.androlib.com/appstats.aspx.
5. Balebako, R., Leon, P.G., Almuhimedi, H., Kelley, P.G., Mugan, J., Acquisti, A., Cranor, L.F., & Sadeh, N. (2011). Nudging users towards privacy on mobile devices. *In Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion.*
6. Boyles, J.L., Smith, A., Madden, M. (2012). Privacy and data management on mobile devices. *Pew Internet & American Life Project.* Retrieved Dec. 31, 2012, from http://pewinternet.org/Reports/2012/Mobile-Privacy/Key-Findings.aspx?view=all.
7. Camerer, C., Issacharoff, S., Samuel, C., & Loewenstein, G. (2003). Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review, 151*, 1211–1254.
8. Chin, E., Felt, A., Sekar, V., & Wagner, D. (2012). Measuring user confidence in smartphone security and privacy. *In Proc. SOUPS 2012*.
9. Cranor, L.F., Guduru, P., & Arjula, M. (2006). User interfaces for privacy agents. *ACM TOCHI, 13*(2), 135–178.
10. Druckman, J. (2001). Using credible advice to overcome framing effects. *J. Law, Economics, and Organization, 17*, 62–82.
11. Egelman, S., Felt, A.P., & Wagner, D. (2012). Choice architecture and smartphone privacy: there's a price for that. *In Proc. WEIS 2012.*
12. Felt, A., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012). Android permissions: user attention, comprehension, and behavior. *In Proc. SOUPS 2012*.
13. Johnson, E.J., Bellman, S., & Lohse, G.L. (2002). Defaults, framing and privacy: why opting in-opting out. *Marketing Letters*, *13*(1), 5–15.
14. Kalnikaitė, V., Rogers, Y., Bird, J., Villar, N., Bachour, K., Payne, S., Todd, P.M., Schöning, J., & Krüger, A. (2011). How to nudge in situ: designing lambent devices to deliver information salience in supermarkets. *In Proc. UbiComp 2011*, 11–20.
15. Kelley, P.G., Cesca, L., Bresee, J., & Cranor, L.F. (2010). Standardizing privacy notices: an online study of the nutrition label approach. *In Proc. CHI 2010*, 1573–1582.

16. Kelley, P.G., Consolvo, S., Cranor, L., Jung, J., Sadeh, N., & Wetherall, D. (2012). A conundrum of permissions: installing apps on an android smartphone. *In Proc. FCDS 2012*, 68–97.
17. Lee, M.K., Kiesler, S., & Forlizzi, J. (2011). Mining behavioral economics to design persuasive technology for healthy choices. *In Proc. CHI 2011,* 325–334.
18. Levin, I.P., & Gaeth, G.J. (1988). Framing of attribute information before and after consuming the product. *J. Consumer Research, 15,* 374–378.
19. Levin, I.P., Schneider, S.L., & Gaeth, G.J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational behavior and human decision processes, 76*(2), 149–188.
20. Lin, J., Amini, S., Hong, J., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. *In Proc. UbiComp 2012*, 501–510.
21. Marteau, T.M. (1989). Framing of information: its influence upon decisions of doctors and patients. *British Journal of Social Psychology, 28,* 89–94.
22. Miller, P.M., & Fagley, N.S. (1991). The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin, 17*, 517–522.
23. MSN Movies. (2012). X-Men. Retrieved in Dec. 31, 2012, from http://movies.msn.com/movies/movie/x-men.
24. Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: the distinction between affective and informational negativity effects. *European Review of Social Psychology, 1,* 33–60.
25. Rogers, Y., Hazlewood, W.R., Marshall, P., Dalton, N., & Hertrich, S. (2010). Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? *In Proc. UbiComp 2010,* 261–270.
26. Taylor, S.E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin, 110,* 67–85.
27. Thaler, R.H., & Sunstein, C.R. (2008). *Nudge: improving decisions about health, wealth, and happiness.* New Haven: Yale University Press.
28. The Wall Street Journal. (2012). What they know. Retrieved Dec. 31, 2012, from http://blogs.wsj.com/wtk-mobile.
29. Tsai, S., Egelman, L., Cranor, L.F., & Acquisti, A. (2007). The effect of online privacy information on purchasing behavior: an experimental study. *In Proc. WEIS 2007.*
30. Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.
31. Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *J. Business, 59*, 251–278.
32. Wired. (2012). Apple says grabbing address book data is an iOS policy violation. Retrieved Dec. 31, 2012, from http://www.wired.com/gadgetlab/2012/02/apple-responds-to-path.