Article:

Okan, Y orcid.org/0000-0001-7963-1363, Stone, ER and Bruine de Bruin, W orcid.org/0000-0002-1601-789X (2018) Designing Graphs that Promote Both Risk Understanding and Behavior Change. Risk Analysis, 38 (5). pp. 929-946. ISSN 0272-4332

https://doi.org/10.1111/risa.12895

Manuscript accepted for publication in the journal Risk Analysis

Designing graphs that promote both risk understanding and behavior change

Yasmina Okan,[1] Eric R. Stone,[2] Wändi Bruine de Bruin [1,3]

[1]Centre for Decision Research, Leeds University Business School, University of Leeds, UK

[2] Department of Psychology, Wake Forest University, US

[3] Department of Engineering and Public Policy, Carnegie Mellon University, US

Correspondence concerning this paper should be addressed to Yasmina Okan, Centre for Decision Research, Leeds University Business School, University of Leeds, Charles Thackrah Building, Leeds, LS2 9LB, United Kingdom; E-mail: y.okan@leeds.ac.uk; Phone: +44 113 343 2622. This study has been funded by a grant awarded by the Worldwide Universities Network (Fund for International Research Collaborations) to Yasmina Okan. Yasmina Okan's time was supported in part by a Population Research Fellowship awarded by Cancer Research UK. Wändi Bruine de Bruin was supported in part by a grant from the Swedish Foundation for the Humanities and the Social Sciences (Riksbankens Jubileumsfond) Program on Science and Proven Experience. The authors declare independence from these funding agencies in each of the following: design of the study; collection, management, analysis, and interpretation of the data; and preparation of the manuscript.

Abstract

Graphs show promise for improving communications about different types of risks, including health risks, financial risks, and climate risks. However, graph designs that are effective at meeting one important risk communication goal (promoting risk-avoidant behaviors) can at the same time compromise another key goal (improving risk understanding). We developed and tested simple bar graphs aimed at accomplishing these two goals simultaneously. We manipulated two design features in graphs, namely whether graphs depicted the number of people affected by a risk and those at risk of harm (‘foreground+background’) vs. only those affected (‘foreground-only’), and the presence vs. absence of simple numerical labels above bars. Foreground-only displays were associated with larger risk perceptions and risk-avoidant behavior (i.e., willingness to take a drug for heart attack prevention) than foreground+background displays, regardless of the presence of labels. Foreground-only graphs also hindered risk understanding when labels were not present. However, the presence of labels significantly improved understanding, eliminating the detrimental effect of foreground-only displays. Labels also led to more positive user evaluations of the graphs, but did not affect risk-avoidant behavior. Using process modelling we identified mediators (risk perceptions, understanding, user evaluations) that explained the effect of display type on risk-avoidant behavior. Our findings contribute new evidence to the graph design literature: Unlike what was previously feared, we demonstrate that it is possible to design foreground-only graphs that promote intentions for behavior change without a detrimental effect on risk understanding. Implications for the design of graphical risk communications and decision support are discussed.

Keywords: Graph design, medical decision making, risk communication

1. INTRODUCTION

In today's information age, people are regulary faced with risk information, including in medical results, commercial advertisements, and the news. For example, an accurate understanding of information about the risks and benefits associated with different medical treatments, screenings, or lifestyle choices is essential for informed medical decision making.[1] Graphical displays can represent risk information in accessible ways, often helping to overcome widespread difficulties in understanding risks and probabilities (e.g., [2–4]). Graphs also constitute effective tools for reducing common judgment biases,[5] changing attitudes,[6] and promoting risk-avoidant behaviors.[7,8] Accordingly, graphs are increasingly being used and recommended as "best practices" for communicating information about people's well-being, such as health risks,[1,9,10] financial risks,[11,12] and climate hazards.[13,14]

However, one serious issue has not yet been addressed: graph designs that are effective at meeting one important risk communication goal can at the same time compromise another key goal. Specifically, graph designs that are effective at improving risk understanding are not well suited to promote risk-avoidant behaviors. This can be an issue in situations in which a key goal of risk communications is to induce people to be more risk averse, e.g., communications designed to motivate people to stop smoking, drive more safely, do more physical exercise, or take steps to prevent sexually transmitted diseases. Conversely, graph designs that are effective at promoting risk-avoidant behaviors can undermine people's risk understanding by making the risks appear larger than they really are, thus hindering people's ability to make informed decisions about their own health.

The existence of a conflict between the goals of increasing risk understanding and promoting risk-avoidant behaviors has been pointed out by several authors.[9,15–17] However, to date, no resolution to this conflict has been found. As noted by Stone et al.,[16] this can pose a problem for risk communicators who aim to achieve both communication goals simultaneously. In the present work we developed and tested simple graph design features aimed at simultaneously accomplishing the goals of promoting risk understanding and risk-avoidant behavior. We also aimed to shed light on the mechanisms underlying any effects of graph design features on risk avoidance. To achieve these aims, we built on the literature on graphical risk communication by considering insights from the cognitive science of graph comprehension.

1.1 Foreground+background vs. foreground-only graphical displays

Displays that depict both the number of people affected by a risk (the numerator of the risk ratio, i.e., the foreground) and the number of people at risk of harm (the denominator of the risk ratio, i.e., the background) are commonly referred to as 'foreground+background' graphs.[(15,16,18)] A common example is stacked bar graphs, in which different segments represent the number of people affected vs. not affected by a risk. Other examples include icon arrays, where icons of different colors represent those affected and not affected by the risk (e.g., with black vs. white icons, respectively), and pie charts. In contrast, 'foreground-only' graphs only depict the number of people affected by a risk. Examples include simple bar graphs representing only those affected by the risk, as well as icon displays where the icons only depict the affected individuals.

There is increasing evidence that foreground+background graphical displays are effective for improving risk understanding, thus facilitating informed decision making (e.g.,[(4,5,19–22)]). To illustrate, Garcia-Retamero and Galesic[(19)] found that people's understanding of the extent to which a new drug reduced the risk of suffering a stroke or heart attack significantly improved with foreground+background graphs (including icon arrays and bar graphs) as compared to numerical information. Foreground+background graphs depict part-to-whole relationships visually. This can help one to disentangle classes that are overlapping in ratios and to bring attention to denominators, which may otherwise be neglected.[(5,9,23,24)]

On the other hand, communicators often aim to promote risk-avoidant behavior. Yet, foreground+background graphs typically result in decreased risk perceptions relative to numerical information, which in turn leads to lower willingness to implement risk-avoidant behavior.[3,18,22] For example, Stone et al.[18] found that participants who received foreground+background graphs in the form of stacked bar charts depicting the reduction in the risk of gum disease associated with a new toothpaste perceived less risk and were less willing to purchase the toothpaste, as compared to those who received the same information in a purely numerical format.

Thus, one concern with the use of foreground+background graphical displays is that they may lead people to dismiss low-probability events, even when public health experts recommend that people take them seriously and implement risk-avoidant behaviors.[16,25] One solution to this problem is to use a foreground-only graphical display to alter the salience of the foreground relative to the background, shifting people's attention to the number of people affected by the risk. When risk magnitudes are small, the focus on affected individuals can lead to increased risk perceptions, resulting in increased risk avoidance.[16,18] A considerable amount of work supports this supposition.[7,8,15,16,18,26,27] For example, in the study by Stone et al.,[18] participants presented with foreground-only graphs perceived the risk reduction associated with the toothpaste as larger and were more willing to purchase it than were those presented with foreground+background graphs or a purely numerical display.

Unfortunately, other work suggests that foreground-only graphs decrease risk understanding in comparison to foreground+background graphs,[16,19] and in some cases even in comparison to a purely numerical display.[15,27] The finding that foreground-only graphs negatively affect people's ability to understand relevant risks raises important questions concerning the ethical desirability of such graphs. Thus, it is important to identify suitable design features that can be added to such graphs in order to avoid potentially harmful misunderstandings of risks.

1.2 Processes involved in graph comprehension and the role of numerical labels

Cognitive models of graph comprehension have identified different perceptual and cognitive processes that people use to extract information from displays such as bar charts, as well as principles to facilitate such processes.[28–33] A key process for accurate graph interpretations is the so-called ‘integration phase,’ which involves inferring information from conventional features in graphs (e.g., axes labels, numerical values on scales, legends, or titles) and integrating this information with that inferred from the visual pattern.[29–33] Unfortunately, there is increasing evidence that people often fail to incorporate information from such conventional features in their interpretations.[34–36] Eye-tracking data suggest that one of the reasons for this failure is that people do not allocate sufficient attention to such features.[36]

In foreground-only graphs, information about the number of people at risk is generally provided in features such as graph titles (e.g.,[18]), legends (e.g.,[16]), or accompanying textual information (e.g.,[19,37]). The fact that people often fail to attend sufficiently to these features may contribute to the detrimental effect of foreground-only graphs on risk understanding. Taking this into account, we designed simple numerical labels aimed at facilitating integration processes and helping people to build an accurate representation of the risk information, regardless of the presence of background information in graphs. The labels contained information about both the foreground and the background (i.e., the number of people who suffered a heart attack and the number of people at risk, respectively) and were positioned immediately above bars in graphs. The positioning of the labels aimed to capitalize on the Gestalt principle of proximity, which suggests that the relative distance of elements to one another determines how these are grouped.[38] Indeed, graphs that make an appropriate use of Gestalt principles of perceptual organization have been shown to facilitate the processes of relating visual patterns to variables and their values.[28,30,32,34,39,40]

1.3 Overview of the present experiment

The main goal of the present study was to test the effectiveness of different graph design features to promote risk understanding as well as risk-avoidant behavior. To this end, we presented participants with graphical information concerning the effectiveness of a hypothetical drug to reduce the risk of suffering a heart attack. We assessed their willingness to take the drug (i.e., risk avoidance) and understanding of the information presented. We also assessed their risk perceptions and evaluations of the graphs. We manipulated two design features in graphs, namely (1) display type (i.e., foreground-only vs. foreground+background), and (2) the presence vs. absence of labels, which displayed the number of people who suffered a heart attack and the total number of people at risk (see Figure 1).

We investigated five research questions concerning the effect of display type and labels, outlined below:

(1) Does the use of a foreground-only display increase willingness to take the drug, and does this advantage remain when labels are added? In line with previous research,[16,18,26] we expected that foreground-only graphs would lead to greater willingness to take the drug than foreground+background displays. We expected that this would be the case regardless of the presence of labels, because labels do not alter the visual salience of the foreground relative to the background.[18]

(2) Does the use of a foreground-only display increase risk perceptions, and does this tendency remain when labels are added? We expected that foreground-only graphs would also increase risk perceptions independently of labels, in line with our rationale for willingness to take the drug. As noted earlier, foreground-only graphs displays can shift

people's attention to the number of people affected by the risk, leading to increased risk perceptions.(7,8,15,16,18,26,27)

(3) Does the use of a foreground-only display decrease risk understanding, and does this problem go away when adding labels? In keeping with previous research,(15,19) we expected that foreground-only displays would be associated with worse risk understanding than foreground+background displays in the absence of labels. In contrast, we predicted that differences in risk understanding associated with the type of display should be reduced or eliminated for graphs containing labels, as these should facilitate graph processing and risk understanding.

(4) What are the effects of display type and presence of labels on user evaluations of the graphs? Assessing user evaluations is relevant because people may not be motivated to attend to, or take actions regarding, graphs that they dislike.(9,16,37) We did not have specific hypotheses concerning user evaluations, as there are no clear a priori reasons to predict that the design features manipulated should lead to more positive vs. negative evaluations.

(5) What are the mechanisms underlying any effects of display type and presence of labels on willingness to take the drug? To address this question we used process modelling testing risk perceptions, risk understanding, and user evaluations as mediators of the link between graph design features and willingness to take the drug. We built on recent work by Stone et al. showing that these factors can account for display type effects on risk avoidance.(16)

Finally, an additional goal of the present work was to examine whether individual differences in numeracy (i.e., the ability to understand and manipulate different numerical expressions of probability,(41,42)) might moderate any effects of numerical labels. People with lower numeracy tend to be less capable of using numerical information when forming their perceptions of risks and benefits,[43,44] or interpreting health risk information.[45] Accordingly, it is possible that any beneficial effect of numerical labels on risk understanding might hold only for more numerate individuals.

2. METHOD

2.1. Participants

Participants were recruited via Amazon's Mechanical Turk (www.mturk.com), which provides access to a paid internet participant panel that has been widely used for behavioral decision-making research.[46,47] The task was available only to individuals who had an acceptance rate greater than or equal to 95% in previous Human Intelligence Tasks (HITs) on Mechanical Turk, following recommendations to ensure high quality data.[48] A total of 1,119 United States residents completed the study. Three participants were removed from subsequent analyses based on a-priori criteria to exclude participants who did not complete the survey in one sitting. The final sample included 1,116 participants. Table I includes the demographic characteristics of the participants. One participant did not provide demographic details.

<Insert Table I about here>

2.2. Materials and design

2.2.1. Medical risk reduction scenario

Participants were asked to imagine that they had symptoms of heart disease and that without taking medication they could have a heart attack. They were further asked to imagine that their risk of having a heart attack would be reduced by taking medication X. Participants then read about a study involving patients with symptoms of heart disease, in which a group of 500 patients had not taken medication X, and another group of 500 patients had taken medication X. The number of patients who had a heart attack in each of the groups was displayed in a bar graph immediately below the introductory text. The bar graph was presented in one of four different formats, as described below. In all cases, the treatment was associated with a 50% relative risk reduction (from 40 affected patients in 500 to 20 affected patients in 500, i.e., a reduction from 8% to 4%). Appendix A includes screenshots of the information viewed by participants in all conditions.

2.2.2. Graphical displays

We manipulated two design features in our bar graphs, namely (1) whether the display type presented a foreground-only vs. foreground+background graph, and (2) whether numerical labels accompanied the bars or not (see Figure 1). In foreground+background graphs, the maximum value on the y-axis scale was the total number of people at risk of a heart attack in each of the groups, such that the y-axis scale ranged from 0 to 500 and increased in increments of 20 (see Figures 1a and 1c). In foreground-only graphs the y-axis scale ranged from 0 to 48 and increased in increments of 2 (see Figures 1b and 1d). In conditions with numerical labels, the labels were displayed immediately above the bars and contained information concerning both the foreground and the background (see Figures 1c and 1d). Such labels were not included in conditions without labels (see Figures 1a and 1b). As noted earlier, the introductory text presented above the graphs contained information about the background in all cases (see Appendix A).

<Insert Figure 1 about here>

2.2.3. Measurement of numeracy

We administered the three numeracy items developed by Schwartz et al.[(49)] and the four items from the Berlin Numeracy Test,[(50)] following recommendations for Mechanical Turk samples.[(50)] For each participant, a composite numeracy score reflected the total number of correct responses across these items (see [(51)] for a similar procedure). Table I provides details on the scores obtained by participants.

2.3. Outcome measures

Appendix B includes all questions corresponding to each of the outcome measures, which are described below. Results for all individual items are provided in the online Supplementary Materials.

2.3.1. Willingness to take the drug

Participants were asked to imagine that the cost of taking medication X would be covered by their insurance and that no other medications were available in the market. Participants then indicated the chance that they would take medication X if it had no side effects, as well as if it had nine different types of side effects, ranging in severity (e.g., stomach cramps, heartburn, gastrointestinal bleeding). An example item is: "What is the chance that you would take medication X if the only common side effect of the medication was stomach cramps?" In all cases, participants gave their response on a scale from 1 to 7, where 1 represented "definitely would not" and 7 represented "definitely would." Responses to the ten items were combined by computing the average score across items (Cronbach's alpha = .91).

2.3.2. Risk perceptions

Risk perceptions were assessed with five items adapted from previous research[18,22] and evaluated participants' opinions concerning the magnitude of the risk of having a heart attack, and of the reduction in the risk achieved by the medicine. Specifically, the first two items assessed the perceived decrease in the risk of having a heart attack resulting from taking medication X ("In your opinion, the decrease in the risk of having a heart attack resulting from taking medication X is…") and the perceived risk reduction significance ("Please indicate how significant you believe the reduction of the risk is when going from taking no medication to taking medication X"), based on Stone et al.[18] The third item asked participants to indicate how helpful they thought the treatment was in reducing the risk of having a heart attack, based on Galesic et al.[22] Finally, the fourth and fifth items asked participants to indicate what they thought was the risk of having a heart attack without medication X and with medication X, respectively (see Appendix B, for more details). The response scale for all items ranged from 1 to 7. The five items were averaged to produce an overall measure of perceived risk (Cronbach's alpha = .78).[1]

2.3.3. Risk understanding

We used eight items to evaluate the accuracy of people's risk understanding, including the accuracy of recall of the risk information presented, and of inferences or operations made on the basis of this information (see[15,20,23]). The first two items assessed whether participants could accurately state the number of people who had suffered a heart attack in the group of people who did not take medication X and in the group of people who took medication X, respectively. Thus, these two items assessed participants' recall of the foreground information, independently of the background. The third and fourth items asked participants to indicate how many of 1,000 patients who have symptoms of heart disease would have a heart attack if they do not take the medication and if they take the medication, respectively. Thus, these items assessed participants' ability to infer what the foreground would be for a different background.

The fifth and sixth items were based on work by Cuite et al.[52] and evaluated participants' ability to perform mathematical operations on the presented risk information. The first of these items asked participants to imagine that there was an alternative medication available, and that 30 out of 1,000 patients who took this medication had a heart attack. Participants then indicated whether they thought medication X or the alternative was more effective. The second item asked participants to indicate how many of 1,000 patients would have a heart attack if an improved version of medication X cut the risk of heart attack in half. Finally, the seventh and eighth items were multiple choice questions adapted from Stone et al.'s approach of presenting a range of options to better capture the gist of people's knowledge.[15] Participants indicated how many of 1,000 people would have a heart attack if they did not take the drug vs. if they took the drug by selecting one of the following options: "a. 0"; "b. 10-20 people"; "c. 30-60 people"; "d. 70-100 people"; "e. more than 100 people". Responses to all items were scored for accuracy (1=correct; 0=incorrect). We computed a total risk understanding score for each participant, out of a total of 8 (Cronbach's alpha = .86).[2]

2.3.4. User evaluations

Participants rated how well they understood the information represented in the graph, how much they liked the way the graph was designed, how much they would trust information represented in a graph like the one they viewed, and how helpful the graph was for deciding about whether or not to take medication X (see [37] for a similar procedure). The response scale for these evaluation questions ranged from 1 to 7. Items were averaged to produce an overall measure of user evaluations (Cronbach's alpha = .84).

2.4. Procedure

The online study was hosted on the web survey platform Unipark (www.unipark.de). Participants read an online informed consent form before proceeding to the study. Subsequently, they were presented with the medical scenario. They were then randomly assigned to one of our four bar graph displays (Figure 1) and answered items corresponding to the different outcome measures, in the following order: willingness to take the drug, risk perceptions, risk understanding, user evaluations, and numeracy.[3] Willingness to take the drug and risk perceptions were assessed one item at a time, while the graph remained visible (see Appendix A). As in previous research,[15,16] items assessing risk understanding and user evaluations were displayed on separate screens, with the graph no longer visible. Doing so prevented participants from just copying answers to the risk understanding questions directly from the bar-graph displays, in items where this was possible (e.g., the first and second understanding items described above). Finally, participants completed basic demographic questions (gender, age, educational level, occupation, nationality, and mother tongue). Time to read the scenario and to answer all questions was unlimited.[4] The study was approved by the Research Ethics Committee of the University of Leeds.

2.5. Data analyses overview

To examine our research questions concerning the effects of display type and the presence of labels on participants' willingness to take the drug, risk perceptions, risk understanding, and user evaluations, we conducted analyses of variance (ANOVAs) on each of these outcome measures, including display type (foreground-only vs. foreground+background) and numerical labels (absent vs. present) as factors. Numeracy scores were included in all ANOVAs to examine whether individual differences in this skill moderated any effects of numerical labels. All models included main effects of numeracy as well as two-way and three-way interactions involving numeracy.[5]

Finally, to examine our research question concerning the mechanisms underlying effects of graph design features on willingness to take the drug, we conducted conditional process analyses using the PROCESS macro for SPSS.[(56)] As noted earlier, we tested the indirect effects of graph design features on participants' willingness to take the drug via risk perceptions, risk understanding, and user evaluations, building on recent process modeling work by Stone et al.[(16)] We used a bias corrected bootstrap procedure based on 5,000 samples, and an indirect effect was considered significant if the 95% confidence interval excluded 0.

3. RESULTS

3.1. Does the use of a foreground-only display increase willingness to take the drug, and does this advantage remain when labels are added?

As expected, participants presented with foreground-only graphs expressed a greater willingness to take the drug, as compared to those presented with foreground+background graphs, F (1, 1108) = 9.46, p = .002, η_p^2 = .01 (Table II). In addition, higher numeracy scores were associated with a greater willingness to take the drug, F (1, 1108) = 8.10, p = .005, η_p^2 = .01. All other effects were not statistically significant (Fs < 1.7, ps > .20), implying that, as expected, the presence of labels did not affect willingness to take the drug.

<Insert Table II about here>

3.2. Does the use of a foreground-only display increase risk perceptions, and does this tendency remain when labels are added?

As expected, risk perceptions were higher in the foreground-only conditions than in the foreground+background conditions, F (1, 1108) = 55.53, p < .001, η_p^2 = .05. Additionally, risk perceptions were generally lower when labels were present than when they were absent, F (1, 1108) = 33.48, p < .001, η_p^2 = .03. There was also an interaction between display type and presence of labels, F (1, 1108) = 5.55, p = .02, η_p^2 = .01. The difference in risk perceptions between foreground-only and foreground+background graphs was smaller for graphs with labels than for graphs without labels, but remained significant in both cases (p = .001, d = .29, and p < .001, d = .59, respectively), as can be seen in Table II.

Finally, the ANOVA also revealed a main effect of numeracy whereby the more numerate provided lower risk perceptions, F (1, 1108) = 24.86, p < .001, η_p^2 = .02. All other effects were not significant (Fs < 2.3, ps > .13).

3.3. Does the use of a foreground-only display decrease risk understanding, and does this problem go away when adding labels?

Overall, foreground+background displays were associated with better risk understanding than foreground-only displays, $F\ (1, 1108) = 12.24$, $p < .001$, $\eta_p^2 = .01$. In addition, the presence of labels improved risk understanding, $F\ (1, 1108) = 157.43$, $p < .001$, $\eta_p^2 = .12$. Most importantly, there was an interaction between display type and presence of labels, $F\ (1, 1108) = 4.81$, $p = .03$, $\eta_p^2 = .004$. Simple effect tests showed that the difference in risk understanding between foreground-only and foreground+background graphs occurred when graphs did not contain labels ($p < .001$, $d = .32$) but not when graphs contained labels ($p = .70$, $d = .04$), as can be seen in Table II. These results support our predictions, suggesting that the detrimental effect of foreground-only displays on risk understanding was eliminated when graphs contained simple numerical labels.

Analyses also revealed a main effect of numeracy whereby higher numeracy was associated with better risk understanding, $F\ (1, 1108) = 203.12$, $p < .001$, $\eta_p^2 = .16$, and an interaction between numeracy and labels, $F\ (1, 1108) = 9.85$, $p = .002$, $\eta_p^2 = .01$. Interestingly, the beneficial effect of labels was stronger among less numerate individuals. The conditional effect of labels on understanding for such individuals (i.e., mean numeracy score – 1 SD) was 2.10 ($p < .001$), whereas among more numerate ones (i.e., mean score + 1 SD) it was 1.21 ($p < .001$). All other effects were not significant ($Fs < 1.3$, $ps > .27$).

3.4. What are the effects of display type and presence of labels on user evaluations of the graphs?

Graphs that included labels received more positive user evaluations than graphs without labels, $F\ (1, 1108) = 19.12$, $p < .001$, $\eta_p^2 = .02$ (see Table II). In addition, higher numeracy was associated with more positive evaluations of the graphs, $F\ (1, 1108) = 8.85$, $p = .003$, $\eta_p^2 = .01$. The ANOVA also revealed an interaction between numeracy, display type, and presence of labels, F (1, 1108) = 4.64, p = .03, η_p^2 = .004. No other significant effects were observed for this outcome measure (Fs < 3.1, ps > .08).[6]

3.5. What are the mechanisms underlying effects of display type and presence of labels on willingness to take the drug?

Table III shows the zero-order correlations among the outcome measures. As can be seen, willingness to take the drug showed a moderately strong positive correlation with risk perceptions, as well as a moderately weak but significant positive correlation with user evaluations. Risk understanding was not correlated with willingness to take the drug. As planned, we examined indirect effects of graph design features on willingness to take the drug via risk perceptions, risk understanding, and user evaluations. Specifically, we built a moderated mediation model with display type as the independent variable, willingness to take the drug as the dependent variable, and risk perceptions, risk understanding, and user evaluations as mediators. Presence of labels was included as a moderator of the effects of display type on each of the mediators tested (see Figure 2).

<Insert Table III about here>

For both the no labels and labels conditions, foreground-only graphs increased risk perceptions, which led to an increased willingness to take the drug. This mediation path was stronger in the no labels condition (conditional indirect effect = .27, within the 95% confidence interval [.19, .36]) than in the labels condition (13, [.06, .21]), as seen in a significant index of moderated mediation –.14 [–.25, –.04]. Yet, as noted earlier there was no interaction between display type and labels on willingness to take the drug (see section 3.1). Hence, we need to additionally consider the link through risk understanding. When no labels were provided, foreground-only displays reduced understanding (see Figure 2). At the same time, understanding was positively associated with willingness to take the drug, and thus the conditional indirect effect of understanding in the no labels condition was to decrease willingness to take the drug (−.02, [−.06, − .01]). However, when labels were provided, display type no longer affected understanding, and thus there was no indirect effect on willingness to take the drug via understanding (−.003, [−.02, .01]). Accordingly, the indirect effect via risk understanding was also moderated by labels (index of moderated mediation = .02 [.004, .06]).

Thus, foreground-only displays consistently increased willingness to take the drug regardless of whether labels were provided, albeit for different reasons. When labels were not provided, such displays significantly increased risk perceptions, which in turn increased willingness to take the drug; however, this effect was tempored by a decrease in risk understanding. When labels were provided, there was a significant but smaller effect via risk perceptions, but this effect was no longer tempored by a decrease in risk understanding. Finally, there was no effect of type of display on willingness to take the drug via user evaluations, both for conditions with and without labels.

<Insert Figure 2 about here>

Next, we built another model that was identical to the previous one, with the exception that presence of labels was included as the independent variable, and display type was included as the moderator. As can be seen in Figure 3, there were multiple significant links with labels despite the lack of a main effect of labels on willingness to take the drug. Specifically, both user evaluations and risk understanding were positively related to willingness to take the drug, and labels increased both of them. The conditional indirect effects via evaluations were 04, [.01, .08] for foreground-only displays and .03, [.01, .07] for foreground+background displays, whereas for risk understanding these were .05, [.0003, .10] for foreground-only displays, and .03, [.001, .07] for foreground+background displays. In contrast, labels reduced risk perceptions, leading to a lower willingness to take the drug. This indirect effect was significant for both types of displays but larger for foreground-only displays than for foreground+background displays (−.23, [.−.31, −.15] and −.08, [.−.16, −.003], respectively). The existence of indirect effects with opposite signs combined with the absence of a significant direct effect on willingness to take the drug (see Figure 3) can explain why labels did not affect this outcome.

<Insert Figure 3 about here>

4. DISCUSSION

In the current work, we built on the graph design literature with the aim of developing a graphical display that could both improve risk understanding and promote behavior change, rather than accomplishing only one of those two goals, as has been the case with previous displays. Building on previous research[(16,18,26)] and in line with our hypotheses, we found that foreground-only graphs were associated with larger risk perceptions and willingness to take a hypothetical drug for heart attack prevention, as compared to foreground+background graphs. Foreground-only displays also resulted in worse risk understanding than foreground+background displays, when graphs did not contain labels. However, our study also provided new evidence that the inclusion of simple numerical labels on top of bars significantly increased risk understanding, eliminating the differences in understanding linked to the type of display. Thus, our findings show that, unlike what was previously feared, [(9,15–17)] it is possible to design foreground-only graphs that promote intentions for behavior change without a detrimental effect on risk understanding.

4.1. Implications for the design of graphical risk communications

Our results are the first to suggest that risk communicators who wish to promote risk-avoidant behaviors without hindering people’s risk understanding may consider the use of foreground-only graphs including numerical labels. These labels should be placed next to the bars to capitalize on Gestalt principles of perceptual organization and facilitate integration processes.[(28,34,40)] Foreground-only displays with labels can encourage people to take steps to reduce risks that threaten their well-being, without limiting their ability to make informed decisions based on an accurate risk understanding. Moreover, we found that graphs with labels were associated with more positive user evaluations, suggesting that this simple design feature is not only useful in improving risk understanding, but also contributes to a positive user experience.

Importantly, our findings indicate that the effectiveness of simple numerical labels is not confined to people who are more skilled with numbers. In our study the beneficial effect of numerical labels was in fact stronger among less numerate participants than among more numerate ones. This unexpected result can be interpreted in the light of previous work showing that lower numeracy is associated with a larger susceptibility to framing and formatting effects (e.g., whether numerical information is presented in a frequency or a percentage format [(57)]).[7]

Our findings also have implications for the development of principles of effective graph design. A key principle put forward by Edward Tufte[(58)] in his theory of graphic design states that graphs should maximize the 'data-ink ratio' (i.e., the proportion of a graph's ink devoted to the non-redundant display of data-information). According to Tufte, one way to accomplish this is to eliminate redundant data-ink (e.g., depicting a given value through the height of bars, the height of the shading in bars, the position of the top horizontal lines, as well as a number placed above bars, all of which would be regarded as redundant design features). However, our results suggest that the inclusion of redundant elements in graphs (e.g., simple numerical labels containing values that can also be inferred by checking the y-axis) may in some cases confer benefits and improve risk understanding. Thus, maximizing the data-ink ratio should not necessarily be the ultimate goal when designing graphs (for similar arguments, see[(59)]). The inclusion of some redundant elements may be particularly useful for drawing attention to information that people do not always focus on, such as axes scales.[36] More generally, our findings converge with recent work showing that insights from the literature on the cognitive science of graph comprehension can help to improve graphical risk communications.[21]

4.2. Mechanisms underlying effects of graph design features

The current work also sheds light on the mechanisms underlying effects of graph design features on behavioral intentions. Foreground-only displays led to larger risk perceptions, which in turn resulted in a greater willingness to take the drug. This finding is consistent with foreground:background salience theory,[18] which suggests that one of the key ways by which graphs influence decision making is through risk perceptions (see also[16,27]). More specifically, this theory suggests that foreground-only graphs are more effective at promoting risk-avoidant behaviors because they call attention to the number of people who are affected, and away from the number of people at risk. It is likely that participants in foreground-only conditions in our study focused to a larger extent on the number of people who suffered a heart attack and to a lesser extent on the total number of people at risk, as compared to those in foreground+background conditions. It is also plausible that participants used the size of bars as a cue for risk magnitude judgments,[23, 35] leading to larger risk perceptions in conditions with larger bars (i.e., foreground-only displays). Indeed, people's perceptions of numerical magnitude can be determined by physical features of graphs including the size of graph elements and the distance between them. For example, distances between two data points along an axis can determine the perceived numerical difference between them.[60,61] More generally, the physical distance between two elements on a plane (e.g., two numbers representing the original price of an item and its discounted price) can determine perceptions of relative numerical difference (e.g., the perceived magnitude of the discount[62]).

Our results also point to an additional pathway underlying the effect of display type on behavioral intentions, namely an effect via risk understanding. In our study, adding labels to graphs reduced risk perceptions in the foreground-only condition, an effect that we had not anticipated. Yet, despite this change in risk perceptions, participants' willingness to take the drug was unaffected by labels. The key to understanding why this occurred is seen in our process modelling. When controlling for the other variables in the model, risk understanding was positively associated with willingness to take the drug. Thus, the improvement in understanding in the foreground-only condition due to labels essentially cancelled the decrease in risk perceptions, leaving the effect on willingness to take the drug unchanged. The effect of graph design on behavioral intentions via risk understanding is not anticipated by foreground:background salience theory. Hence, our findings are in agreement with other recent research suggesting that foreground:background salience does not provide a sufficient explanation of participants' behavioral intentions,[63] and that future theoretical developments should include other determinants of behavior, in this case, risk understanding.

An interesting question that arises from our work is why the positive (albeit small) link between understanding and willingness to take the drug occurred. One possibility is that this relationship occurred due to increased self-efficacy (i.e., the perceived belief concerning one's capability to execute a given behavior). Better understanding can result in increased self-efficacy, which is known to be a key determinant of health behavior[64,65].[8] Moreover, our results also revealed a positive link between user evaluations of the graphs and behavioral intentions. It is plausible that self-efficacy is enhanced when people are presented with communication materials that they like, trust, and perceive to be helpful, resulting in a greater willingness to act (see also[16,37]).

Finally, another interesting question is why labels reduced risk perceptions. Although labels should not alter the visual salience of the foreground relative to the background, it is likely that the numerical information provided called attention to both elements, thus reducing risk perceptions. The finding that the negative link between labels and risk perceptions was stronger for foreground-only graphs than for foreground+background graphs supports this interpretation, as the latter type of graphs already displayed both the foreground and the background graphically.

4.3. Limitations and future research

As with all studies, the current work has limitations. First, the fact that our study involved a sample of Mechanical Turk users suggests some caution regarding immediate prescriptive applications of our findings to the entire population. In general, the educational attainment level of Mechanical Turk users is higher than that of the US population.(47) Moreover, women tend to be overrepresented in US Mechanical Turk samples,(66,67) as also reflected in our sample (see Table 1). Additionally, our finding that the beneficial effect of numerical labels was weaker among more numerate participants may reflect to some extent a ceiling effect limiting the impact of labels among such individuals. Thus, future research should test whether findings generalize to more diverse samples.

Second, we measured participants' behavioral intentions in one hypothetical setting. Future work should examine whether our findings generalize to actual health behavior, as well as to other behaviors that can affect people's well-being (e.g., financial behavior). Finally, the current study involved relatively low probabilities (8% vs. 4%). Depicting the background graphically for low-probability risks visually shows that the probability is low, thereby reducing perceived risk and risk avoidance.(9,16,18,23,26) However, if the probabilities are higher, adding the background will likely not have the same effect, since the probability depicted will be larger. Preliminary results support this claim, pointing to key boundary conditions of the effects documented here.(68)

4.4. Conclusions

Previous research in graphical risk communication has recognized a tension between the goals of improving risk understanding and promoting healthy behaviors. Recommendations have therefore focused on addressing each goal separately. Although supporting informed decision making about health is undoubtedly an essential risk communication goal, promoting risk-averse behaviors can often constitute an important and ethically desirable goal in public health. Our findings suggest that interventions designed to promote risk avoidance (e.g.,foreground-only graphs displaying the number of people affected by a risk) do not necessarily need to come at a cost of reduced risk understanding.

Footnotes

[1] The fifth item (which asked about the risk of having a heart attack with medication X) had an item-total correlation of .20, and deleting it improved Cronbach's alpha to .84. The same main effects and interactions reported below were found for composite measures of risk perceptions excluding vs. including this item. These four-item and five-item composite measures were highly correlated ($r = .97$, $p < .001$). Here we report results including all five items.

[2] The item-total correlation for item 5 (i.e., the item asking participants to compare the effectiveness of medication X with that of a different medication) was .25, and deleting it improved Cronbach's alpha to .88. The same main effects and interactions reported below were found for composite measures of understanding excluding vs. including this item. These seven-item and eight-item composite measures were highly correlated ($r = .98$, $p < .001$). Here we report results including all eight items.

[3] We also measured participants' need for cognition (i.e., the tendency for an individual to engage in and enjoy thinking[(53)]), need for cognition for numerical items in particular, using an adapted scale developed by the current authors, and graph literacy (i.e., the ability to understand graphically presented information[(54)]). These scales were included between the user evaluation items and the numeracy scale. All main effects and interactions reported below remained unchanged when need for cognition and graph literacy scores were included as covariates in analyses, with the exception of the main effect of numeracy for user evaluations, which was no longer statistically significant ($p = .27$).

[4] The average completion time across the entire study, including all individual differences scales, was 25 minutes. Considering recent recommendations for detecting inattention in online studies,[(55)] we computed the 5% trimmed mean completion time (24 min. 51 s.), and reran our analyses excluding the participants who completed the study in less than half of this

time (n= 45). All main effects and interactions reported below remained unchanged and thus all analyses reported include the full sample.

[5] To obtain independent effects, we standardized numeracy scores by converting them to z-scores prior to computing the interaction terms. Main effects and interactions involving numeracy were included as covariates in ANOVAs.

[6] Among low numerate participants, the beneficial effect of labels on evaluations was stronger for foreground-only displays, whereas among high numerate participants this effect was stronger for foreground+background displays. As we have no theoretical basis for why this 3-way interaction occurred and have tested a number of interactions with numeracy across the dependent measures, it may just reflect a Type 1 error.

[7] We thank an anonymous reviewer for offering this insight.

[8] Interestingly, there was no zero-order correlation between risk understanding and willingness to take the drug (see Table III), but the partial correlation between these variables after controlling for risk perceptions was .11 ($p < .001$). The lack of a zero-order correlation may reflect two opposite processes occurring simultaneously. On the one hand, people with poor understanding may have reduced self-efficacy, leading to a reduced willingness to take the drug. On the other hand, people with poor understanding may view the risk as greater than it actually is, leading to increased willingness to take the drug. This possibility would explain why the understanding-willingness to take the drug link only becomes significant when controlling for risk perceptions. Future research should measure perceived self-efficacy to more directly test this possibility.

REFERENCES

1. Garcia-Retamero R, Cokely ET. Communicating health risks with visual aids. Current Directions in Psychological Science. 2013; 22:392–399.
2. Petrova D, Garcia-Retamero R, Cokely ET. Understanding the harms and benefits of cancer screening: A model of factors that shape informed decision making. Medical Decision Making. 2015; 35:847–858.
3. Waters EA, Weinstein ND, Colditz GA, Emmons KM. Reducing aversion to side effects in preventive medical treatment decisions. Journal of Experimental Psychology: Applied. 2007; 13:11–21.
4. Zikmund-Fisher B, Ubel P, Smith D, Derry H, McClure J, Stark A, et al. Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. Patient Education and Counseling. 2008; 73:209–14.
5. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. Journal of Behavioral Decision Making. 2012; 25:390–401.
6. Garcia-Retamero R, Cokely ET. Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. Journal of Experimental Psychology: Applied. 2011; 17:270–287.
7. Schirillo JA, Stone ER. The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. Risk Analysis. 2005; 25:555–566.
8. Stone ER, Yates JF, Parker AM. Effects of numerical and graphical displays on professed risk-taking behavior. Journal of Experimental Psychology: Applied. 1997; 3:243–256.

9. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: A systematic review. Journal of the American Medical Informatics Association. 2006; 13:608–618.

10. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. Medical Decision Making. 2007; 27:695–713.

11. Duclos R. The psychology of investment behavior: (De)biasing financial decision-making one graph at a time. Journal of Consumer Psychology. 2015; 25:317–325.

12. Raghubir P, Das SR. The long and short of it: Why are stocks with shorter runs preferred? Journal of Consumer Research. 2010; 36:964–982.

13. Lorenz S, Dessai S, Forster PM, Paavola J. Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. Philosophical Transactions of the Royal Society A. 2015; 373:20140457.

14. Taylor AL, Dessai S, de Bruin WB. Communicating uncertainty in seasonal and interannual climate forecasts in Europe. Philosophical Transactions of the Royal Society A. 2015; 373:20140454.

15. Stone ER, Gabard AR, Groves AE, Lipkus IM. Effects of numerical versus foreground-only icon displays on understanding of risk magnitudes. Journal of Health Communication. 2015; 20:1230–1241.

16. Stone ER, Bruine de Bruin W, Wilkins AM, Boker EM, MacDonald Gibson J. Designing graphs to communicate risks: Understanding how the choice of graphical format influences decision making. Risk Analysis. 2017; 37:612–628.

17. Brewer N. Goals. In: Fischhoff B, Brewer N, Downs JS, editors. Communicating risks and benefits: An evidence-based user's guide. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2012. p. 3–10.

18. Stone ER, W. R Sieck, Bull BE, Yates JF, Parks SC, Rush CJ. Foreground: background salience: Explaining the effects of graphical displays on risk avoidance. Organizational Behavior and Human Decision Processes. 2003; 90:19–36.

19. Garcia-Retamero R, Galesic M. Who profits from visual aids: Overcoming challenges in people's understanding of risks. Social Science & Medicine. 2010; 70:1019–1025.

20. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Education and Counseling. 2008; 73:448–455.

21. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. Journal of Experimental Psychology: Applied. 2015; 21:178–194.

22. Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: Overcoming low numeracy. Health Psychology. 2009; 28:210–216.

23. Reyna VF. A theory of medical decision making and health: Fuzzy trace theory. Medical Decision Making. 2008; 28:850–865.

24. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learning and Individual Differences. 2008; 18:89–107.

25. Shepperd JA, Lipkus IM, Sanderson SC, McBride CM, O'Neill SC, Docherty S. Testing different communication formats on responses to imagined risk of having versus missing the GSTM1 gene. Journal of Health Communication. 2013; 18:124–137.

26. Hu T-Y, Jiang X-W, Xie X, Ma X-Q, Xu C. Foreground-background salience effect in traffic risk communication. Judgment and Decision Making. 2014; 9:83–89.

27. Chua HF, Yates JF, Shah P. Risk avoidance: Graphs versus numbers. Memory & Cognition. 2006; 34:399–410.

28. Ali N, Peebles D. The effect of Gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2013; 55:183–203.

29. Carpenter PA, Shah P. A model of the perceptual and conceptual processes in graph comprehension. Journal of Experimental Psychology: Applied. 1998; 4:75–100.

30. Kosslyn SM. Understanding charts and graphs. Applied Cognitive Psychology. 1989; 3:185–226.

31. Lohse GL. A cognitive model for understanding graphical perception. Human-Computer Interaction. 1993; 8:353–388.

32. Pinker S. A theory of graph comprehension. In: Freedle R, editor. Artificial intelligence and the future of testing. Hillsdale, NJ: Erlbaum; 1990. p. 73–126.

33. Shah P, Carpenter PA. Conceptual limitations in comprehending line graphs. Journal of Experimental Psychology: General. 1995; 124:43–61.

34. Huestegge L, Philipp AM. Effects of spatial compatibility on integration processes in graph comprehension. Attention, Perception, & Psychophysics. 2011; 73:1903–1915.

35. Okan Y, Garcia-Retamero R, Galesic M, Cokely ET. When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. Spatial Cognition & Computation. 2012; 12:1–25.

36. Okan Y, Galesic M, Garcia-Retamero R. How people with low and high graph literacy process health graphs: Evidence from eye-tracking. Journal of Behavioral Decision Making. 2016; 29:271–294.

37. Bruine de Bruin W, Stone ER, Gibson JM, Fischbeck PS, Shoraka MB. The effect of communication design and recipients' numeracy on responses to UXO risk. Journal of Risk Research. 2013; 16:981–1004.

38. Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, et al. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. Psychological Bulletin. 2012; 138:1172–1217.

39. Peebles D, Ali N. Differences in comprehensibility between three-variable bar and line graphs. In: Proceedings of the thirty-first annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum Associates.; 2009. p. 2938–2943.

40. Shah P, Mayer RE, Hegarty M. Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. Journal of Educational Psychology. 1999; 91:690–702.

41. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical Decision Making. 2001; 21:37–44.

42. Peters E. Beyond comprehension: The role of numeracy in judgments and decisions. Current Directions in Psychological Science. 2012; 21:31–35.

43. Dieckmann NF, Slovic P, Peters EM. The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. Risk Analysis. 2009; 29:1473–1488.

44. Lipkus IM, Peters E, Kimmick G, Liotcheva V, Marcom P. Breast cancer patients' treatment expectations after exposure to the decision aid program adjuvant online: The influence of numeracy. Medical Decision Making. 2010; 30:464–473.

45. Gardner PH, McMillan B, Raynor DK, Woolf E, Knapp P. The effect of numeracy on the comprehension of information about medicines in users of a patient information website. Patient Education and Counseling. 2011; 83:398–403.

46. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. Annual Review of Clinical Psychology. 2016; 12:53–81.

47. Paolacci G, Chandler J. Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science. 2014; 23:184–188.

48. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behavior Research Methods. 2014; 46:1023–1031.

49. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Annals of Internal Medicine. 1997; 127:966–972.

50. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: The Berlin Numeracy Test. Judgment and Decision Making. 2012; 7:25–47.

51. Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated samples: Numeracy and metacognition. Judgment and Decision Making. 2014; 9:15–34.

52. Cuite CL, Weinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Medical Decision Making. 2008; 28:377−384.

53. Cacioppo JT, Petty RE. The need for cognition. Journal of Personality and Social Psychology. 1982; 42:116–131.

54. Galesic M, Garcia-Retamero R. Graph literacy: A cross-cultural comparison. Medical Decision Making. 2011; 31:444–457.

55. Maniaci MR, Rogge RD. Caring about carelessness: Participant inattention and its effects on research. Journal of Research in Personality. 2014; 48:61–83.

56. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford Press; 2013.

57. Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychological Science. 2006; 17:407–413.

58. Tufte ER. The visual display of quantitative information. 2nd ed. Cheshire, CT: Graphics Press; 2001.

59. Stephen Few. The chartjunk debate. A Close examination of recent findings. Visual Business Intelligence Newsletter [Internet]. 2011 [cited 2016 Dec 15]; Available from:

https://perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

60. Sun Y, Li S, Bonini N. Attribute salience in graphical representations affects evaluation. Judgment and Decision Making. 2010; 5:151–158.
61. Sun Y, Li S, Bonini N, Su Y. Graph-framing effects in decision making. Journal of Behavioral Decision Making. 2012; 25:491–501.
62. Coulter KS, Norberg PA. The effects of physical distance between regular sales prices on numerical difference perceptions. Journal of Consumer Psychology. 2009; 19:144–157.
63. Stone ER, Reeder EC, Parillo J, Long, C, Walb, L. Salience versus proportional reasoning: Rethinking the mechanism behind graphical display effects. Journal of Behavioral Decision Making. Under review.
64. Bandura A. Health promotion by social cognitive means. Health Education & Behavior. 2004; 31:143–64.
65. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implementation Science. 2012; 7.
66. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods. 2012; 44:1–23.
67. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judgment and Decision Making. 2010; 5:411–419.
68. Parillo J, Stone E, Okan Y, Bruine de Bruin W, Parker A. When context matters: The impact of different probability sizes and risk reductions on graphical display effects. Poster presented at the 37th Annual Meeting of the Society for Judgment and Decision Making. 2016 Nov; Boston, MA.

Table I

Participant characteristics

Participant characteristics	
Gender	
Male	373 (33.5%)
Female	742 (66.5%)
Age	
Range	18 – 82
Mean	37.3
Standard deviation	12.6
Education	
High school or less	105 (9.4%)
Some college or associate degree	412 (37.0%)
Bachelor’s degree	406 (36.4%)
Master’s degree or higher	192 (17.2%)
Occupation	
High school student	4 (.4%)
Apprentice/trainee	5 (.4%)
College student	95 (8.5%)
Employed	723 (64.8%)
Unemployed	225 (20.2%)
Retiree	63 (5.7%)
Nationality	
US	1095 (98.2%)
Other	20 (1.8%)
English as a first language	1093 (98.0%)
Numeracy	
Range	0 – 7
Mean	3.3
Standard deviation	1.7

Table II

Mean scores for the outcome measures, as a function of presence vs. absence of labels and display type (SEM in parentheses).

	Labels absent		Labels present	
	Foreground-only	Foreground+background	Foreground-only	Foreground+background
Willingness to take drug (1-7)	4.37_a (.07)	4.18_b (.07)	4.38_a (.07)	4.14_b (.07)
Risk perception (1-7)	4.70_a (.06)	4.11_b (.06)	4.22_b (.06)	3.94_c (.06)
Risk understanding (0-8)	4.24_a (.17)	5.11_b (.16)	6.24_c (.12)	6.33_c (.12)
User evaluations (1-7)	5.42_a (.07)	5.37_a (.08)	5.73_b (.06)	5.64_b (.07)

Note: Non-matching subscripts in each row indicate significant pairwise contrasts at $p < .05$.

Table III

Zero-order correlations between the outcome measures.

	Risk perception	Risk understanding	User evaluations
Willingness to take drug	.42 ***	.00	.22 ***
Risk perception		-.23 ***	.18 ***
Risk understanding			.25 ***

Note: *** $p < .001$

Figure 1

Graphs viewed by participants in the different experimental conditions. (A) Foreground+background, without labels; (B) Foreground-only, without labels; (C) Foreground+background, with labels; (D) Foreground-only with labels.

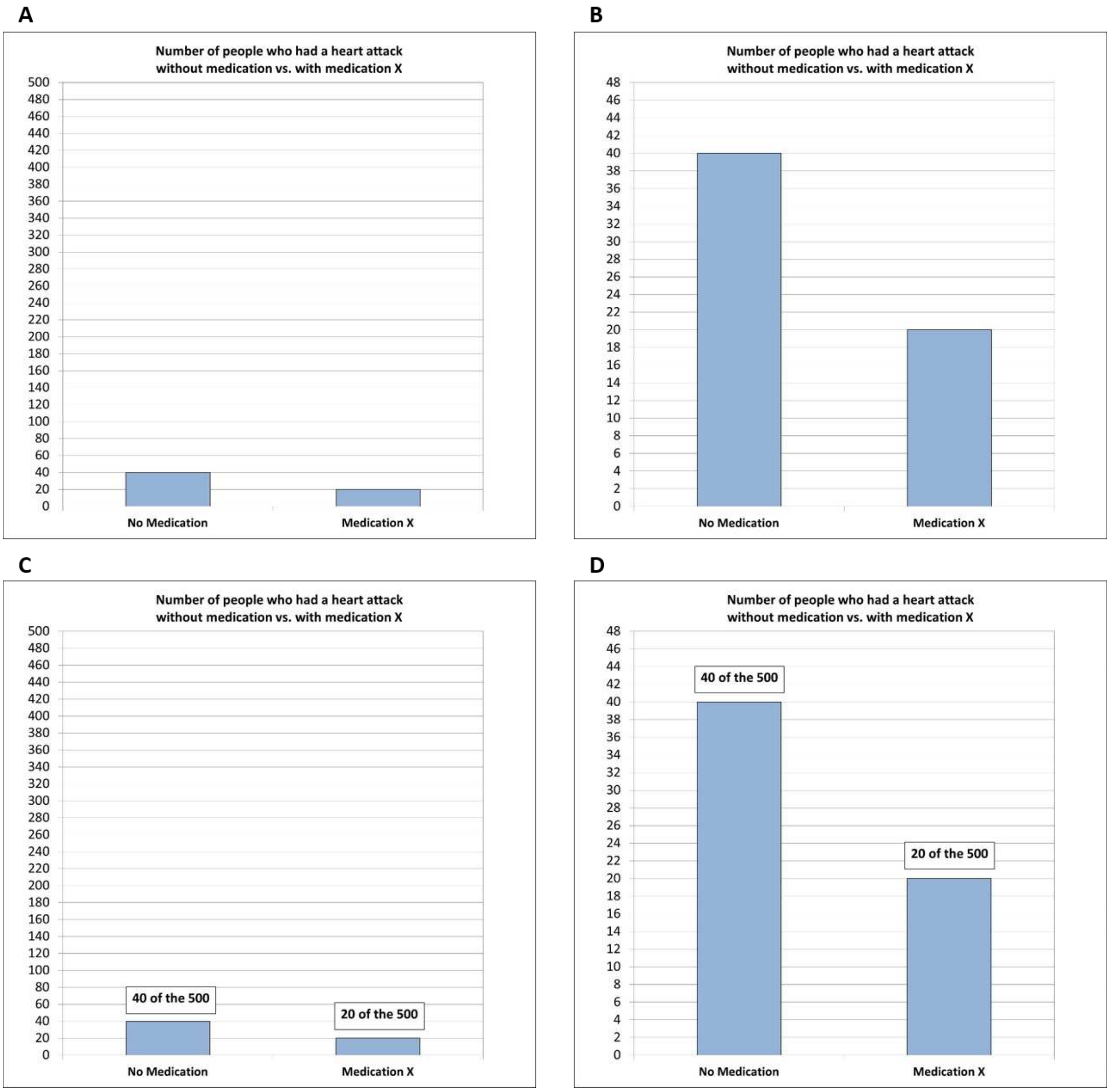


Figure 2

Conditional process analyses of the effect of display type on willingness to take the drug. Coefficients for the paths between display type and each of the mediators (risk understanding, risk perception, and evaluations) represent conditional effects with a simple effects parameterization. The path between display type and willingness to take the drug represents the direct effect (unconditional), after controlling for the mediators. Note: Unstandardized coefficients are shown. The scale for risk perception, evaluations, and willingness to take the drug was from 1 to 7, whereas for risk understanding it was from 0 to 8. †p < .10, *p ≤ .05, **p ≤ .01, *** p ≤ .001.

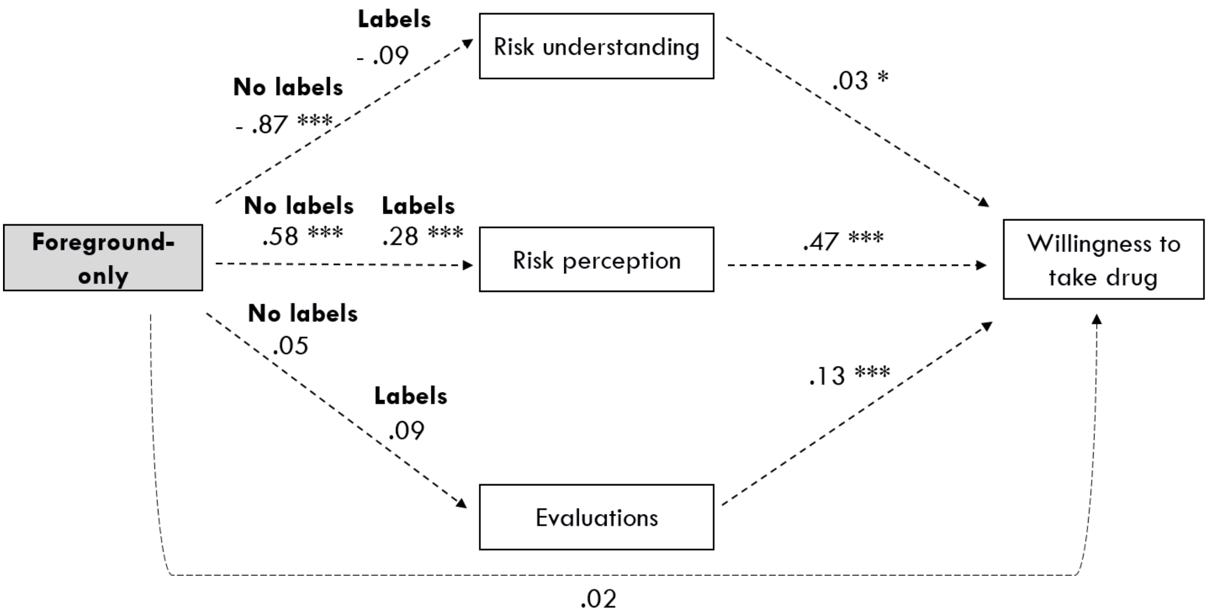
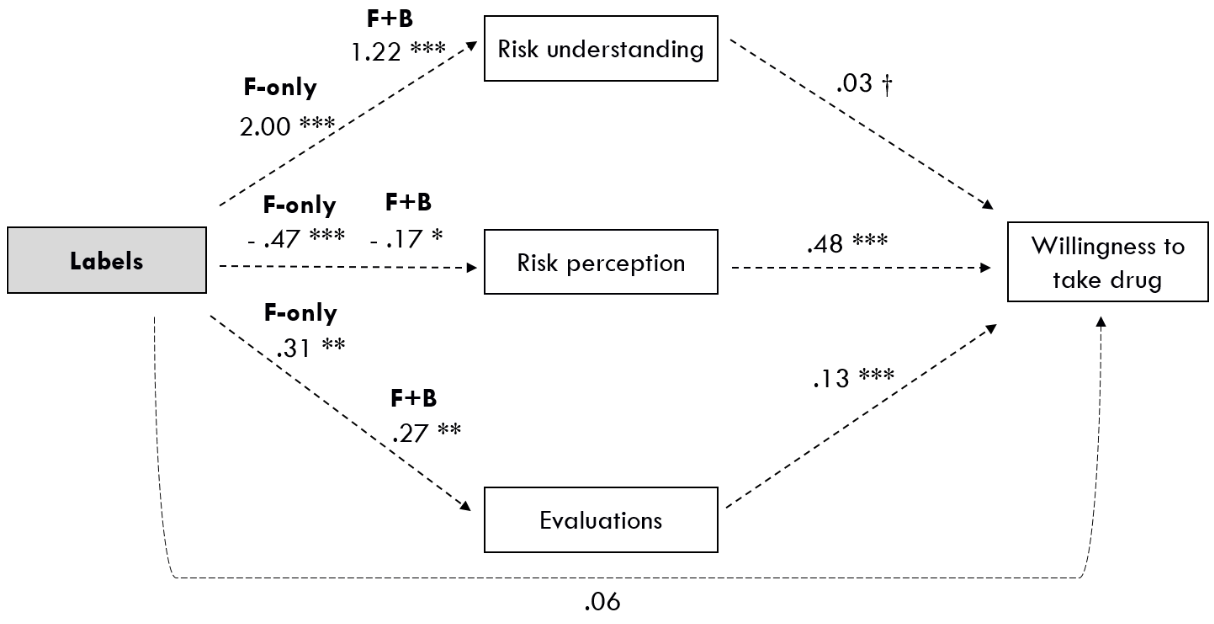


Figure 3

Conditional process analyses of the effect of labels on willingness to take the drug (F-only = foreground only displays; F+B = foreground+background displays). Coefficients for the paths between labels and each of the mediators (risk understanding, risk perception, and evaluations) represent conditional effects with a simple effects parameterization. The path between labels and willingness to take the drug represents the direct effect (unconditional), after controlling for the mediators. Note: Unstandardized coefficients are shown. The scale for risk perception, evaluations, and willingness to take the drug was from 1 to 7, whereas for risk understanding it was from 0 to 8. †p < .10, *$p \leq .05$, **$p \leq .01$, *** $p \leq .001$.



APPENDIX A

Screenshots of scenario presented to participants in all conditions. The question displayed corresponds to the first item assessing willingness to take the drug.

Figure A1. Foreground+background, no labels

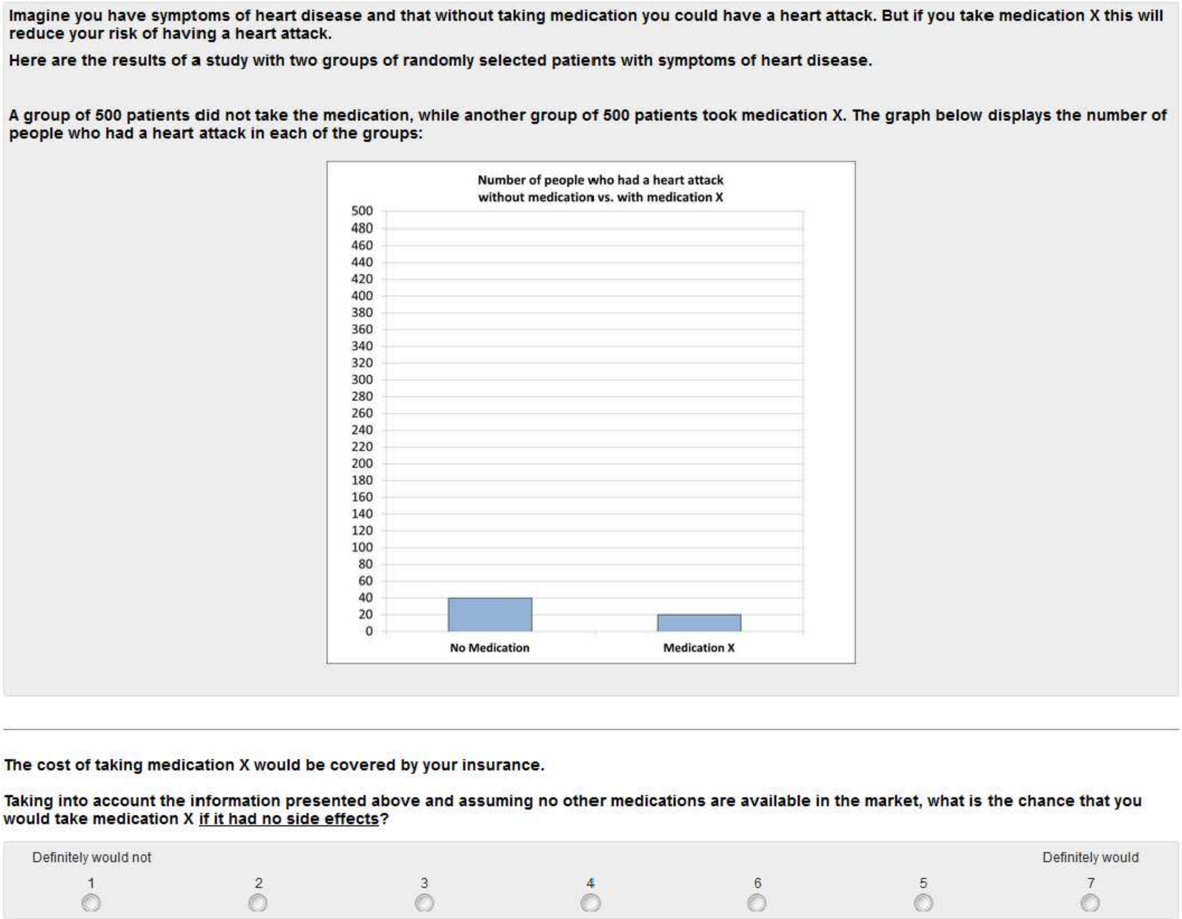
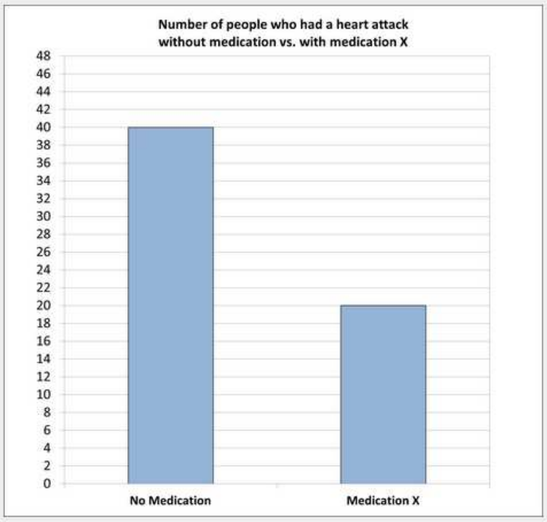


Figure A2. Foreground-only, no labels

Imagine you have symptoms of heart disease and that without taking medication you could have a heart attack. But if you take medication X this will reduce your risk of having a heart attack.

Here are the results of a study with two groups of randomly selected patients with symptoms of heart disease.

A group of 500 patients did not take the medication, while another group of 500 patients took medication X. The graph below displays the number of people who had a heart attack in each of the groups:



The cost of taking medication X would be covered by your insurance.

Taking into account the information presented above and assuming no other medications are available in the market, what is the chance that you would take medication X <u>if had no side effects</u>?

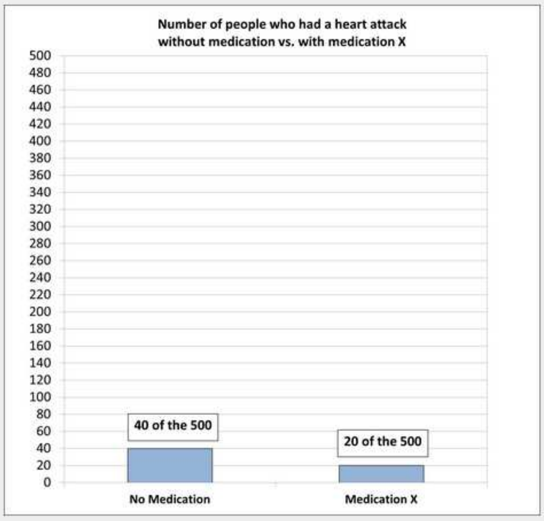
Definitely would not						Definitely would
1	2	3	4	6	5	7

Figure A3. Foreground+background, labels

Imagine you have symptoms of heart disease and that without taking medication you could have a heart attack. But if you take medication X this will reduce your risk of having a heart attack.

Here are the results of a study with two groups of randomly selected patients with symptoms of heart disease.

A group of 500 patients did not take the medication, while another group of 500 patients took medication X. The graph below displays the number of people who had a heart attack in each of the groups:



The cost of taking medication X would be covered by your insurance.

Taking into account the information presented above and assuming no other medications are available in the market, what is the chance that you would take medication X <u>if it had no side effects</u>?

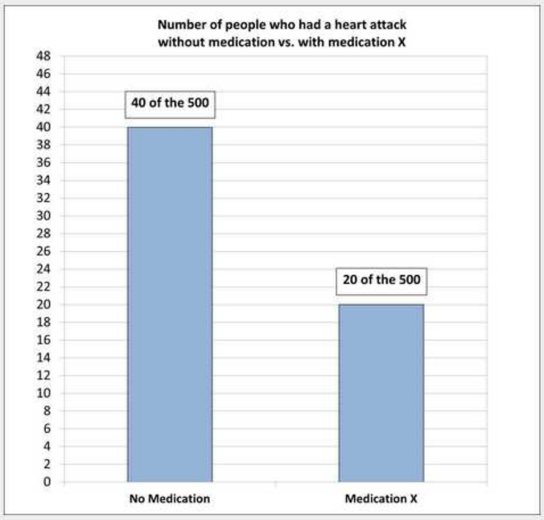
Definitely would not						Definitely would
1	2	3	4	6	5	7
○	○	○	○	○	○	○

Figure A4. Foreground-only, labels

Imagine you have symptoms of heart disease and that without taking medication you could have a heart attack. But if you take medication X this will reduce your risk of having a heart attack.

Here are the results of a study with two groups of randomly selected patients with symptoms of heart disease.

A group of 500 patients did not take the medication, while another group of 500 patients took medication X. The graph below displays the number of people who had a heart attack in each of the groups:



The cost of taking medication X would be covered by your insurance.

Taking into account the information presented above and assuming no other medications are available in the market, what is the chance that you would take medication X <u>if it had no side effects?</u>

Definitely would not						Definitely would
1	2	3	4	6	5	7
○	○	○	○	○	○	○

APPENDIX B

Willingness to take the drug

The cost of taking medication X would be covered by your insurance.
Taking into account the information presented above and assuming no other medications are available in the market, what is the chance that you would take medication X

1. if it had no side effects?

if the only common side effect for the medication was…

2. Stomach cramps
3. Heartburn
4. Headache
5. Loss of appetite
6. Throwing up
7. Hives and swelling
8. Ringing in the ears
9. Gastrointestinal bleeding
10. Haemorrhage within the skull

Note: In all cases, 1 = definitely would not, 7 = definitely would

Risk perceptions

1. In your opinion, the decrease in the risk of having a heart attack resulting from taking medication X is… (1= None, 7 = incredibly big)

2. Please indicate how significant you believe the reduction of the risk is when going from taking no medication to taking medication X by selecting a number on the scale below. In other words: When I compare the amount of risk of having a heart attack associated with taking no medication to the risk of having a heart attack associated with medication X, I'd say the difference is. . . (1= insignificant, 7 = highly significant)

3. How helpful do you think medication X was in reducing the risk of having a heart attack? (1= not at all helpful, 7 = very helpful)

4. What do you think is the risk of having a heart attack <u>without medication X</u>? (1= extremely low, 7 = extremely high)

5. What do you think is the risk of having a heart attack <u>with medication X</u>? (1= extremely low, 7 = extremely high)

Risk understanding

1. How many people had a heart attack in the group of patients who took no medication? (correct = 40)

2. How many people had a heart attack in the group of patients who took medication X? (correct = 20)

3. Imagine 1000 patients who have symptoms of heart disease. If they take no medication, how many of them will have a heart attack?

   ___________ out of 1,000 (correct = 80)

4. Imagine 1000 patients who have symptoms of heart disease. If they take medication X, how many of them will have a heart attack?

   ___________ out of 1,000 (correct = 40)

5. Imagine that another medication is available in the market (medication Y). Assume that a group of 1000 patients take medication Y, and 30 of them have a heart attack.
   Which medication is more effective at preventing heart attack?

   a. Medication X, b. Medication Y, c. They are equally effective (correct = b)

6. Imagine that future research has led to medication X being improved, so that the risk of having a heart attack with medication X is cut in half.
   Out of 1000 patients who take the medication, how many would you expect to have a heart attack?

   ___________ out of 1,000 (correct = 20)

7. Out of every 1000 people who take no medication, how many will have a heart attack?

   a. 0; b. 10-20 people; c. 30-60 people; d. 70-100 people; e. more than 100 people (correct = d)

8. Out of every 1000 people who take medication X, how many will have a heart attack?

   a. 0; b. 10-20 people; c. 30-60 people; d. 70-100 people; e. more than 100 people (correct = c)

User evaluations

1. How well did you understand the information represented in the graph? (1 = not at all, 7 = very well)

2. How much do you like the way the graph was designed? (1 = did not like it at all, 7 = liked it a lot)

3. How much would you trust information represented in a graph like the one you just viewed? (1= not at all, 7 = very much)

4. How helpful is the graph for deciding about whether or not to take medication X? (1= not helpful, 7 = very helpful)