Word count= 6,750

Manuscript accepted for publication in Risk Analysis

Probability size matters: The effect of foreground-only vs. foreground+background graphs on risk aversion diminishes with larger probabilities

Yasmina Okan,[1 *] Eric R. Stone,[2] Jonathan Parillo,[3] Wändi Bruine de Bruin,[1,4] Andrew M. Parker [5]

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

[2]Department of Psychology, Wake Forest University, Winston-Salem, NC, USA

[3]Department of Psychology, Vanderbilt University, Nashville, TN, USA

[4]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA

[5]RAND Corporation, Pittsburgh, PA, USA

* Address correspondence to Yasmina Okan, Centre for Decision Research, Leeds University Business School, University of Leeds, Charles Thackrah Building, Leeds, LS2 9LB, UK, Phone: +44 113 343 2622; E-mail: y.okan@leeds.ac.uk

Acknowledgements

This work was supported by Wake Forest University (US) and by the Worldwide Universities Network (Fund for International Research Collaborations). Jonathan Parillo was based at Wake Forest University at the time the research was conducted. Yasmina Okan's time was in part supported by a Population Research Fellowship awarded by Cancer Research UK (reference C57775/A22182). Wändi Bruine de Bruin was in part supported by a grant from the Swedish Riksbanken Jubilieumsfond programme on Science and Proven Experience. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Abstract

Graphs are increasingly recommended for improving decision making and promoting risk-avoidant behaviors. Graphs that depict only the number of people affected by a risk (‘foreground-only’ displays) tend to increase perceived risk and risk aversion (e.g., willingness to get vaccinated), as compared to graphs that also depict the number of people at risk for harm (‘foreground+background’ displays). However, previous research examining these ‘foreground-only effects’ has focused on relatively low-probability risks (<10%), limiting generalizability to communications about larger risks. In two experiments, we systematically investigated the moderating role of probability size on foreground-only effects, using a wide range of probability sizes (from .1% to 40%). Additionally, we examined the moderating role of the size of the risk reduction, i.e., the extent to which a protective behavior reduces the risk. Across both experiments, foreground-only effects on perceived risk and risk aversion were weaker for larger probabilities. Experiment 2 also revealed that foreground-only effects were weaker for smaller risk reductions, while foreground-only displays decreased understanding of absolute risk magnitudes independently of probability size. These findings suggest that the greater effectiveness of foreground-only versus foreground+background displays for increasing perceived risk and risk aversion diminishes with larger probability sizes and smaller risk reductions. Moreover, if the goal is to promote understanding of absolute risk magnitudes, foreground+background displays should be used rather than foreground-only displays regardless of probability size. Our findings also help to refine and extend existing theoretical accounts of foreground-only effects to situations involving a wide range of probability sizes.

Keywords: Graph design; decision making; risk perception; risk aversion; risk communication

Summary for social media (max. 200 characters)

New research on graphical risk communication finds that the strength of a well-documented display effect (the ‘foreground-only’ effect) is affected substantially by the probability size represented.

1. INTRODUCTION

Effective risk communication is vital for improving public understanding of threats and for informing decisions about potential risk-reduction actions. The question of how best to communicate risk has received a lot of attention (Spiegelhalter, 2017; Trevena et al., 2013; Zipkin et al., 2014). A common recommendation is to use simple graphical displays such as icon arrays, in which icons symbolize the number of individuals who are affected by a risk. Such simple graphs can improve risk understanding (Galesic, Garcia-Retamero, & Gigerenzer, 2009; Garcia-Retamero & Galesic, 2010; Okan, Garcia-Retamero, Cokely, & Maldonado, 2012, 2015; Zikmund-Fisher et al., 2008) and promote risk-avoidant behaviors (Garcia-Retamero & Cokely, 2011, 2014).

The impact of graphs, however, is largely determined by specific design features. ‘Foreground-only’ displays depict graphically only the number of people affected by a risk, whereas the total number of people potentially at risk is provided only numerically (e.g., in the title; see Fig. 1a). In contrast, ‘foreground+background’ displays depict graphically both the number of people affected and the number of people at risk (see Fig. 1b). Recipients tend to see risks as larger when presented with foreground-only displays, as compared to foreground+background displays or purely numerical formats. As a result, foreground-only displays often increase perceived risk and risk aversion (Chua, Yates, & Shah, 2006; Hu, Jiang, Xie, Ma, & Xu, 2014; Okan, Stone, & Bruine de Bruin, 2018; Schirillo & Stone, 2005; Stone et al., 2003; Stone, Bruine de Bruin, Wilkins, Boker, & MacDonald Gibson, 2017; Stone, Gabard, Groves, & Lipkus, 2015; Stone, Yates, & Parker, 1997). Examples of risk aversion include willingness to engage in risk reduction actions such as getting vaccinated or taking a drug to prevent disease, willingness to recommend such actions to others, and willingness to support risk-mitigating policies to increase public knowledge and safety.[1]

A key limitation of previous research on ‘foreground-only effects’ is that it has primarily focused on relatively low-probability risks, typically below 10% (Table I). Yet, risks to people’s wellbeing may sometimes be larger. For instance, the lifetime risk of a disabling injury over 50 years of driving without a seatbelt is about 33% (Slovic, Fischhoff, & Lichtenstein, 1978), and the risk of fatal coronary heart disease among men with risk factors can be as high as 43% (Berry et al., 2012). It is unclear whether foreground-only effects exist beyond the relatively small probability sizes examined in previous research. We aimed to fill this gap in the literature by systematically examining the role of probability size in foreground-only effects.

1.1. Theoretical accounts of the foreground-only effect

Two main theories have been developed to explain foreground-only effects, with a focus on low-probability risks. The *foreground:background salience* theory (Stone et al., 2003) posits that foreground-only displays increase risk aversion because they draw attention to the number of people affected, and away from the overall number of people at risk. For example, the foreground-only display in Fig. 1a draws attention to the 10 people affected (i.e., the foreground), and away from the 1,000 people at risk (i.e., the background). For low-probability risks, this focus on affected individuals increases risk perceptions and in turn risk aversion, as compared to a foreground+background display or a purely numerical one. In contrast, the foreground+background display in Fig. 1b makes both the foreground and the background salient, highlighting that the number of affected individuals is out of many more people at risk, thus reflecting a small risk.

A related theoretical account (*proportional reasoning;* Stone, Reeder, Parillo, Long, & Walb, 2018) states that what matters is whether the graph facilitates forming the proportion, rather than the salience per se. This account builds on the notion that foreground+background displays visually depict the part-to-whole relationship or proportion, whereas foreground-only displays do not (Ancker, Senathirajah, Kukafka, & Starren, 2006). Specifically, Stone et al. (2018) argued that foreground-only displays limit people's ability to engage in proportional reasoning because they display the foreground and the background in different modalities (graphical vs. numerical). For low-probability risk information, difficulties in perceiving the part-to-whole relationship makes it hard to see that the risk is small (Reyna, 2008), leading to greater perceived risk and risk aversion.

Although the precise mechanisms outlined by the two theoretical accounts differ slightly, both accounts suggest that foreground-only displays make it harder to see that low-probability risks are small (by calling attention away from the background and/or by making it difficult to form the proportion) in comparison to foreground+background displays, which make the low-probability nature of the risk "pop out" (Reyna & Brainerd, 2008, p. 103). Such mechanisms can also help explain the finding that foreground-only displays tend to reduce risk understanding, documented in some studies (Stone et al., 2015, 2017). By making it harder to see that risks are small, such displays would lead to overestimations of the risks and hence to worse understanding of absolute risk magnitudes.

Notably, however, the mechanisms outlined by both accounts have only been applied to explain foreground-only effects for low-probability risks. Considering these mechanisms, we reasoned that any overestimation produced by foreground-only graphical displays in relation to foreground+background displays should be reduced for larger probabilities, since the risk magnitudes actually are large. Thus, we expected that foreground-only displays should be less likely to lead to risk overestimation as displayed probability sizes increase. That is, foreground-only effects should be weaker at larger probability sizes.

1.2. Previous studies on the foreground-only effect and the role of probability size and risk reduction

Although previous work has generally focused on probabilities below 10% (Table I), some exceptions exist. One study used probabilities of 9% and 11% (Shepperd et al., 2013) and did not find a foreground-only effect. This is consistent with our claim that foreground-only effects should be weaker at larger probability sizes. The other exception involved probabilities as high as 20% (Garcia-Retamero & Galesic, 2010), but did not assess perceived risk or risk aversion.

In addition to probability size, a second factor that could influence the strength of foreground-only effects is the magnitude of the risk reduction. Previous studies have often included scenarios where a behavior (e.g., getting vaccinated) or characteristic (e.g., not having a genetic marker) reduces a risk (e.g., of disease). However, as seen in Table I, these studies have typically used large relative risk reductions, often around 50% (e.g., a reduction from 8% to 4%). Large risk reductions may look more substantial in foreground-only (vs. foreground+background) displays, increasing the perceived decrease in risk and risk aversion. In contrast, the Shepperd et al. (2013) study that did not replicate the foreground-only effect presented a smaller relative risk reduction of 18%, which might appear trivial regardless of whether it is depicted in a foreground-only or a foreground+background display. Thus, we predicted that the foreground-only effect on perceived decrease in risk and risk aversion would be weaker for smaller risk reductions.

<Table I>

1.3. Present studies and research questions

In two experiments, we tested whether the foreground-only effect varies by probability size. We presented participants with icon arrays depicting the likelihood of contracting a fictitious disease, using either a foreground-only display or a foreground+background display. The displays presented probabilities ranging from the low end examined in previous research (.1%) to probabilities that substantially exceeded those (40%). In the second experiment we also manipulated the risk reduction associated with different vaccines (i.e., how much they reduced the likelihood of contracting the disease).

Our primary research question was:

RQ1: (a) Do foreground-only displays produce greater risk aversion, perceived risk, and perceived decrease in risk than foreground+background displays, and (b) are these effects moderated by probability size and (c) risk reduction level? Following previous research (Hu et al., 2014; Okan et al., 2018; Stone et al., 2003, 2017), we expected that foreground-only displays would increase participants' risk aversion (e.g., willingness to recommend a vaccine to protect against the disease), perceived risk (including cognitive and affective components),[2] and perceived decrease in risk, relative to foreground+background displays. As noted above, we expected that these effects would be weaker at larger probability sizes and smaller risk reductions.

We also addressed two additional research questions:

RQ2: (a) Do foreground-only displays decrease risk understanding relative to foreground+background displays, and (b) is this effect moderated by probability sizes? We expected that foreground-only displays would reduce risk understanding relative to foreground+background displays, in keeping with previous research (Garcia-Retamero & Galesic, 2010; Stone et al., 2015, 2017). Although existing accounts of foreground-only effects suggest that foreground-only displays will lead to reduced understanding (Ancker et al., 2006; Stone et al., 2003, 2018), it is unclear whether or not those mechanisms will hold for increased probability sizes. Thus, we did not generate specific hypotheses regarding the moderating effect of probability size for risk understanding.

RQ3: (a) What is the effect of display type on user evaluations, and (b) is this effect moderated by probability size? We also assessed participants' evaluations of the displays (i.e., liking and perceived usefulness; Ancker et al., 2006; Bruine de Bruin et al., 2013; Okan et al., 2018; Stone et al., 2017). We had no a priori reasons to predict that evaluations of icon arrays would vary by display type, or that any differences in evaluations would be moderated by probability size.

Both experiments were approved by the Institutional Review Board of Wake Forest University. All study materials and data are available from the Open Science Framework (https://doi.org/10.17605/OSF.IO/UYRWF) (Okan, Stone, Parillo, Bruine de Bruin, & Parker, 2019).

2. EXPERIMENT 1

In Experiment 1 we investigated whether foreground-only displays produce greater risk aversion and perceived risk in comparison to foreground+background displays (RQ1a) and whether probability size (from .1% to 40%) moderated these effects (RQ1b). We also examined effects of display type and probability size on risk understanding (RQ2) and user evaluations (RQ3).

2.1. Method

2.1.1. Participants.

Participants were recruited via Amazon's Mechanical Turk (MTurk). We recruited individuals who had an acceptance rate of at least 95% in previous MTurk tasks, following recommendations to ensure quality data (Peer, Vosgerau, & Acquisti, 2014). A total of 2,078 United States residents accessed our study, and 2,006 completed it. Following recommendations for detecting inattention in online studies (Maniaci & Rogge, 2014), we excluded 190 participants who completed the study in less than half of the 5% trimmed mean completion time (1 min 46 s). We also excluded six outliers who took more than two hours. The final sample included 1,810 participants.

2.1.2. Materials and design.

Participants were asked to imagine that they lived in an area that the Centers for Disease Control (CDC) had identified as being affected by the infectious disease "Slibitis." They received information about its symptoms (similar to flu, including fever, sweats, headaches, back pain and physical weakness) and its duration (between 2 weeks and several months).

Additionally, the probability of contracting Slibitis was presented using one of ten icon arrays, varying display type and probability size. Display type was manipulated by depicting either only the number of people affected by Slibitis (foreground-only; Figs. 1a and 1c), or both the number of people affected and the number of people at risk (foreground+background; Figs. 1b and 1d). Probability size was manipulated by varying the number of affected people, either 1, 10, 100, 200, or 400. In all cases the total number of people at risk was 1,000, resulting in probability sizes of, respectively, .1%, 1%, 10%, 20%, or 40%.

All displays included a title stating the total number of people at risk (1,000) and a legend indicating what black and white circles represented (i.e., people affected vs.

not affected by the risk). To maximize any effects of the manipulation of display type, the number of people affected vs. not affected (e.g., 10 vs. 990) was represented only graphically, without accompanying numerical information.

< Fig. 1>

2.1.3. Outcome measures.

In both experiments we used multi-item outcome measures (adapted from Bruine de Bruin et al., 2013; Okan et al., 2018; Stone et al., 2017, 2015). Unless otherwise stated, itemized rating scales included numeric labels for all scale points together with verbal labels for the endpoints (described below). All item-level results for both experiments are presented in Supplementary materials.

2.1.3.1. Risk aversion. First, participants indicated whether they would recommend their friends and family to take a vaccine to protect against Slibitis using a scale from 1 (*definitely no*) to 7 (*definitely yes*). Second, participants indicated what percentage of the budget they would designate towards researching Slibitis (vs. HIV/AIDS) if they worked for the CDC, using a scale from 0% to 100% (in increments of 10%). Third, participants were informed that the CDC presently spends $10,000 on educating the public about Slibitis and indicated whether this amount should be modified using a scale from 1 (*spend a lot less*) to 7 (*spend a lot more*). The three items were z-scored and averaged (Spearman-Brown coefficient=.67).

2.1.3.2. Perceived risk. Participants indicated their perceived chance of contracting Slibitis on a scale from 1 (*extremely low*) to 7 (*extremely high*) and rated how worried they would be about their chances of contracting Slibitis on a scale from 1 (*not at all*) to 7 (*extremely*). The two items were averaged (Spearman-Brown coefficient=.87).

2.1.3.3. Risk understanding. Participants were asked to imagine that they lived in a city of 100,000 people and to indicate how many people in their city they would expect to contract Slibitis: “a. between 0 and 40”; “b. between 50 and 200”; “c. between 500 and 1,000”; “d. between 2,000 and 5,000”; or “e. between 10,000 and 100,000”. Responses were scored for accuracy (1=correct; 0=incorrect).

2.1.3.4. User evaluations. Participants rated how well they understood the likelihood information in the graph, how helpful they found the graph for making decisions about Slibitis, how much they liked how the graph was designed, and how much they would trust information represented in a graph like the one they viewed. The response scale ranged from 1 to 7 in all cases. The four items were averaged (Spearman-Brown coefficient=.86).

2.1.4. Procedure.

The study was conducted online using Qualtrics. After providing informed consent, participants were randomly assigned to an experimental condition and presented with the corresponding display. Participants then completed the first user evaluation item, followed by items assessing risk aversion, perceived risk, risk understanding, and the remaining user evaluation items.[3]

All items were presented after the graph was no longer visible because we were interested in assessing people’s responses based on their mental representations of the information (following Stone et al. 2015, 2017, 2018). Although some studies have allowed participants to look at graphs while completing outcome measures such as risk understanding (e.g., Bruine de Bruin et al., 2013), we removed graphs to mimic situations in which people read information, form a mental representation of it, and then need to make judgments and decisions based on that representation.

2.1.5. Analysis plan.

To address our primary research questions (RQ1a and RQ1b), we conducted analyses of variance (ANOVAs) on risk aversion and perceived risk with display type (foreground-only vs. foreground+background) and probability size (.1%, 1%, 10%, 20%, vs. 40%) as between-subject factors. To facilitate visual inspection of the results, we also constructed graphs of predicted means (estimated using linear regression with probability size and its natural logarithm as predictors, along with 95% confidence intervals) for risk aversion and perceived risk at each probability size condition, by display type condition. To address our research questions about risk understanding (RQ2) and user evaluations (RQ3), we conducted analogous ANOVAs on these outcomes. For all outcomes, ANOVAs yielding significant interactions were followed by independent samples *t*-tests comparing foreground-only and foreground+background displays at each of the probability sizes. In the manuscript we report the main effects and interactions that pertain to our research questions. Full ANOVA findings for both experiments are included in Supplementary materials (Tables S6 and S7).

2.2. Results

2.2.1. RQ1: (a) Do foreground-only displays produce greater risk aversion and perceived risk than foreground+background displays, and (b) are these effects moderated by probability size?

2.2.1.1. Risk aversion. As expected, foreground-only displays (M_z=.06, SD_z=.75) elicited more risk aversion than did foreground+background displays (M_z=-.06, SD_z=0.76), $F(1,1798)=10.91$, $p=.001$, $\eta_p^2=.01$. Supporting our prediction, probability size moderated the effect of display type, $F(4,1798)=5.57$, $p<.001$, $\eta_p^2=.01$. Specifically, there was no significant foreground-only effect for probabilities of 10% and above (Table II).

<Table II>

2.2.1.2. Perceived risk. Participants presented with foreground-only displays (*M*=3.76, *SD*=1.62) perceived Slibitis as riskier than those presented with foreground+background displays (*M*=2.86, *SD*=1.32), $F(1,1800)=206.63$, $p<.001$, $\eta_p^2=.10$. Again, probability size moderated the effect of display type, $F(4,1800)=6.91$, $p < .001$, $\eta_p^2=.02$. Although the foreground-only effect on perceived risk occurred at all probability sizes, its size generally diminished as probability size increased (Table II).

2.2.1.3. Predicted means for risk aversion and perceived risk. Fig. 2 presents predicted means for risk aversion and perceived risk (solid lines, with 95% confidence intervals in dashed lines), for each probability size and display type, with linear interpolation between points within display-type condition. The foreground-only results are shown in blue, and the foreground+background in red. For risk aversion, the two regions start to intersect (and hence the foreground-only effect starts to disappear) between probabilities of 1% and 10% (Fig. 2a). In contrast, for perceived risk it remains through 40% (Fig. 2b).

<Fig. 2>

2.2.2. *RQ2: (a) Do foreground-only displays decrease risk understanding relative to foreground+background displays, and (b) is this effect moderated by probability size?*

There was no main effect of display type on understanding, $F(1,1800)=0.09$, $p=.76$, $\eta_p^2<.001$. However, there was an interaction between display type and probability size, $F(4,1800)=17.50$, $p<.001$, $\eta_p^2=.04$. For .1% probability, the foreground-only display (*M*=.30, *SD*=.46) produced worse understanding than the foreground+background display (*M*=.64, *SD*=.48), $t(353)=6.92$, $p<.001$. In contrast, the foreground-only display produced *better* understanding than the foreground+background display at probabilities of 10% (*M*=.50, *SD*=.50 and *M*=.31, *SD*=.47, respectively), *t*(362)=3.65, *p*<.001, and 20% (*M*=.51, *SD*=.50 and *M*=.34, *SD*=.48, respectively), *t*(362)=3.29, *p*=.001. Differences in understanding between display type for probability sizes of 1% and 40% were not significant (*p*s>.22).

2.2.3. RQ3: (a) What is the effect of display type on user evaluations, and (b) is this effect moderated by probability size?

Participants presented with foreground-only displays (*M*=3.76, *SD*=1.68) gave less positive evaluations than those presented with foreground+background displays (*M*=5.18, *SD*=1.28), *F*(1,1800)=437.89, *p* < .001, η_p^2=.20. Interestingly, probability size moderated this effect, *F*(4,1800)=18.35, *p*<.001, η_p^2=.04. Although foreground-only displays were evaluated less positively at all probability levels (all *p*s<.001), this was particularly pronounced at the .1% probability (*M*=2.97, *SD*=1.70 and *M*=5.54, *SD*=1.22 for the foreground-only and foreground+background conditions, respectively), *t*(353)=16.25, *p*<.001. Rerunning the ANOVA after excluding the .1% probability size yielded a non-significant interaction between display type and probability size, *F*(3, 1447)=1.42, *p*=.23, η_p^2=.003, indicating that the preference for the foreground+background display was consistent across the other probability conditions.

2.3. Discussion

Experiment 1 revealed that foreground-only effects varied substantially by probability size. Participants presented with foreground-only displays depicting small probabilities viewed the risk of contracting a disease as higher than those presented with foreground+background displays. Similar but weaker patterns were observed for larger probabilities, supporting our predictions. Moreover, foreground-only displays resulted in higher risk aversion relative to foreground+background displays, but only for the smallest probabilities sizes (.1% and 1%).

We also found that foreground-only displays were only associated with impaired risk understanding for the smallest probability size (.1%). Contrary to our expectations, larger probability sizes yielded an effect in the opposite direction. This unexpected result is likely an artifact of the response options provided. For probabilities of 10% - 40%, the correct answer was “between 10,000 and 100,000,” which was the largest response option. Thus, respondents overestimating the risk would have answered that question correctly. Experiment 2 included different measures of risk understanding to overcome this limitation.

Finally, foreground-only displays were evaluated less positively than foreground+background displays. This effect was particularly salient at the smallest probability size. In this condition, the foreground-only display depicted the probability as a single circle representing one person affected with the disease, which participants may have found strange or confusing. To clarify that each icon represents one person and to conform with current recommendations (Kreuzmair, Siegrist, & Keller, 2017; Zikmund-Fisher et al., 2014), Experiment 2 replaced circles with stick figures.

3. EXPERIMENT 2

Experiment 2 used two risk reduction scenarios involving vaccines that reduced the likelihood of developing a disease. We investigated whether foreground-only displays produced greater risk aversion, perceived risk, and perceived decrease in risk than foreground+background displays (RQ1a), and whether probability size (from .5% to 40%) moderated these effects (RQ1b). Additionally, we examined the moderating role of risk reduction level (RQ1c). We also addressed our two additional questions about effects of display type and probability size on risk understanding (RQ2) and user evaluations (RQ3).

3.1. Method

3.1.1. Participants.

The recruitment procedure was identical to Experiment 1, except that access to the survey via mobile phones or tablets was not permitted, to optimize the appearance of the graphical displays. A total of 2,145 United States residents accessed our study, and 2,058 completed it. We excluded 126 inattentive participants following a similar procedure as in Experiment 1 (i.e., those who completed the study in less than half of the 5% trimmed mean completion time, namely 2 min 34 s), and one outlier who completed the survey in over two hours.[4] The final sample included 1,931 participants (1,131 women, age range 18-75, *M*=35.76, *SD*=12.10). Ten percent had no more than a high school diploma, 38% had completed up to some college or associate degree, 36% had a bachelor's degree, and 16% had a master's degree or higher.

3.1.2. Materials and design.

The scenario was the same as in Experiment 1, except that participants were informed that two different vaccines had been developed (Vaccines X and Y) to reduce the likelihood of contracting Slibitis. We manipulated the relative risk reduction associated with the vaccines within-subjects, which was set to either 20% or 80% (Table III).

For each vaccine, participants viewed two icon arrays presented side by side. The left icon array depicted the likelihood of contracting Slibitis without any vaccine, and the right with one of the two vaccines (Fig. 3). As in Experiment 1, participants received one of two display types (foreground-only or foreground+background) and one of five probability sizes. The probability sizes without the vaccine were identical to those in Experiment 1 except for the smallest probability, which was set to .5%. This change was made to avoid using fractions of icons in the graphical displays, which would have been necessary in the .1% condition to represent the risk of contracting Slibitis with a vaccine.

<Table III>

< Fig. 3>

3.1.3. Outcome measures.

As in Experiment 1, itemized rating scales included numeric labels for all scale points together with verbal labels for the endpoints, unless otherwise stated.

3.1.3.1. Risk aversion. First, participants indicated on a 1 to 7 scale (1=*definitely no*, 7 =*definitely yes*) whether they would recommend the vaccine to their friends and family, assuming that the vaccine was the only one available for Slibitis. Second, participants were asked to imagine that they went to a local pharmacy and saw that the clinic was offering the vaccine free of charge, and indicated how long they would be willing to wait to get vaccinated. Third, they indicated how much they would be willing to pay for the vaccine. The latter two items included verbal labels for all scale points representing, respectively, timeframes of increasing length (e.g., 1=*would not get the vaccine regardless of the wait time*, 4=*would get the vaccine if the wait was under 15 minutes*, 7=*would get the vaccine regardless of the amount of time I would have to wait*) and increasing amounts of money (1=*$*0, 4=*$20 to $50*, 7=*more than $200*). The three items were averaged within each risk-reduction condition (Spearman-Brown coefficient=.87 in the 20% risk reduction condition; .83 in the 80% risk reduction condition).

3.1.3.2. Perceived risk. Participants assessed the chances of contracting Slibitis (1=*extremely low*, 7=*extremely high*) and how worried they would be about their chances of contracting it without any vaccine (1=*not worried at all*, 7=*extremely worried*). The two items were averaged within each risk-reduction condition (Spearman-Brown coefficient=.86 in the 20% risk reduction condition; .87 in the 80% risk reduction condition).[5]

3.1.3.3. Perceived decrease in risk. Participants assessed the decrease in the likelihood of contracting Slibitis (1=*none*, 7=*incredibly big*) and how much less worried they would be about contracting Slibitis if they received the vaccine (1=*no reduction in worry*, 7=*much less worried*). The two items were averaged within each risk-reduction condition (Spearman-Brown coefficient=.74 in both the 20% and 80% risk reduction conditions).

3.1.3.4. Risk understanding. Participants were asked “about how many people out of 1,000” would contract Slibitis if they do not get a vaccine, if they get Vaccine X, and if they get Vaccine Y. Because previous work has shown that display type can differentially influence understanding of absolute and relative risk magnitudes (Stone et al., 2015), we constructed measures of each. To score *understanding of the absolute risk magnitudes*, we took participants’ estimates of the number of people contracting the disease—without the vaccine, with Vaccine X, and with Vaccine Y—and compared them to the actual absolute magnitudes. We scored responses as correct if within 25% of the correct value so as to better capture how well participants understood the approximate magnitude of the absolute risk (i.e., the gist; Reyna, 2008; Stone et al., 2015) and summed the three scores. For example, when the exact correct value was 10, responses between 7.5 and 12.5 were treated as correct. To score *understanding of the relative risk magnitudes*, we followed established procedures to compute participants' estimates of the relative risk reduction achieved by each vaccine (Garcia-Retamero & Galesic, 2010; Schwartz, Woloshin, Black, & Welch, 1997), as well as their estimates of the risk reduction achieved by one vaccine relative to the other. The three resulting estimates were compared to the actual risk reductions in each case (.20, .80, and .75) and scored as correct or incorrect.[6]

3.1.3.5. User evaluations. We assessed user evaluations as in Experiment 1 (Spearman-Brown coefficient=.80).

3.1.4. Covariates.

Following Larson et al. (2015), participants indicated whether they believed that vaccines can protect from serious diseases (yes/no/maybe). They also answered four items assessing their general attitudes towards vaccines on 1 to 7 scales, which were averaged. Finally, they indicated how long ago they got the last seasonal flu vaccine. Responses to the latter question were categorized into four groups (got the vaccine the past flu season/got the vaccine prior to the past flu season/never got a vaccine/don´t know or don´t remember) and this variable was dummy coded for analyses, using 'never got a vaccine' as the reference group.

3.1.5. Procedure.

Procedures for recruitment and random assignment to experimental conditions were identical to Experiment 1. Information about Vaccine X was presented first, followed by information about Vaccine Y. We counterbalanced whether Vaccine X or Vaccine Y provided the 20% vs. 80% risk reduction. Items assessing risk aversion, perceived risk, and perceived decrease in risk were presented following the information for each of the vaccines. After completing these items for both vaccines, participants completed measures of risk understanding and user evaluations. Finally, they completed the covariate measures and answered demographic questions (gender, educational level, age, and income). As in Experiment 1, all items were presented when the graph was no longer visible.

3.1.6. Analysis plan.

To address our primary research questions (RQ1a, RQ1b, and RQ1c), we conducted analyses of covariance (ANCOVAs) on risk aversion, perceived risk, and perceived decrease in risk, with display type (foreground-only vs. foreground+background) and probability size (.5%, 1%, 10%, 20%, vs. 40%) as between-subject factors and risk reduction level (20% vs. 80%) as a within-subjects factor. Order of presentation (20% risk reduction first vs. 80% risk reduction first) was also included as a between-subjects factor.[7] To better control error variance, we included pre-selected covariates (outlined in section 3.1.4) and demographics. To address our research questions about risk understanding (RQ2) and user evaluations (RQ3), we conducted analogous ANCOVAs on these outcomes, without the within-subject factor of risk-reduction level. For all outcomes, overall analyses yielding significant interactions were followed with ANCOVAs comparing foreground-only and foreground+background displays at each of the probability sizes (and risk reduction level, for analyses including this factor).

3.2. Results

3.2.1. RQ1: (a) Do foreground-only displays produce greater risk aversion, perceived risk, and perceived decrease in risk than foreground+background displays, and (b) are these effects moderated by probability size and (c) risk reduction level?

3.2.1.1. Risk aversion. Participants presented with foreground-only displays (M=3.84, SD=1.52) were more risk averse than those receiving foreground+background displays (M=3.61, SD=1.55), $F(1,1902)$=18.74, p<.001,

η_p^2=.01. As in Experiment 1, the effect of display type was moderated by probability size, *F*(4,1902)=4.59, *p*=.001, η_p^2=.01, with no foreground-only effect for probabilities of 20% or higher (Table IV). As expected, risk reduction level also moderated the effect of display type, *F*(1,1902)=14.47, *p*<.001, η_p^2=.01, with a weaker effect for the 20% vs. the 80% risk reduction level. Specifically, the foreground-only effect was small or nonexistent at all probability levels in the 20% risk reduction condition (Table IV).

<Table IV>

3.2.1.2. Perceived risk. Participants who viewed foreground-only displays (*M*=4.19, *SD*=1.69) perceived the risk without any vaccine as larger than those who viewed foreground+background displays (*M*=3.35, *SD*=1.55), *F*(1,1902)=196.02, *p*<.001, η_p^2=.09. Again, probability size moderated the effect of display type, although the interaction did not reach conventional levels of significance, *F*(4,1902)=2.10, *p*=.08, η_p^2=.004. There was no interaction between display type and risk reduction level, *F*(1,1902)=.43, *p*=.51, η_p^2<.001, which is to be expected given that perceived risk reflects cognitive risk perceptions and worry *without* the vaccine. Thus, the size of the risk reduction achieved by the vaccine should not be relevant.

3.2.1.3. Perceived decrease in risk. Foreground-only displays (*M*=3.80, *SD*=1.43) resulted in a greater perceived decrease in risk than foreground+background displays (*M*=3.46 *SD*=1.50), *F*(1,1902)=46.33, *p*<.001, η_p^2=.02. Again, probability size moderated the effect of display type, *F*(4,1902)=4.30, *p*=.002, η_p^2=.01, with the foreground-only effect being weaker at higher probability sizes. In addition, risk reduction level moderated the effect of display type, *F*(1,1902)=33.16, *p*<.001, η_p^2=.02. Although there were robust foreground-only effects in the 20% risk reduction condition for probabilities of 10% or less, these were smaller than in the 80% risk reduction condition (Table IV).

3.2.1.4. Predicted means for risk aversion, perceived risk, and perceived decrease in risk. Fig. 4 mirrors Fig. 2, with the foreground-only effect represented by the difference between the foreground-only (blue) and foreground+background (red) regions. For risk aversion, the two conditions are relatively indistinguishable across all probabilities in the 20% risk reduction condition (indicating little foreground-only effect), whereas in the 80% risk reduction condition the effect is quite strong at low probabilities, largely disappearing around 20% probability (Fig. 4a). The foreground-only effect for perceived risk also reduces with probability size, but does not completely go away by probability levels of 40% (Fig. 4b). Finally, perceived decrease in risk shows only a small foreground-only effect (and only at small probability sizes) in the 20% risk reduction condition, whereas in the 80% risk reduction condition it shows a prominent foreground-only effect that diminishes with probability size, disappearing between 20% and 30% (Fig. 4c).

<Fig. 4>

3.2.2. RQ2: (a) Do foreground-only displays decrease risk understanding relative to foreground+background displays, and (b) is this effect moderated by probability size?

Foreground-only displays (*M*=1.58, *SD*=1.34) produced worse understanding of absolute risk magnitudes than foreground+background displays (*M*=1.82, *SD*=1.23), $F(1,1902)=23.27$, $p<.001$, $\eta_p^2=.01$. Probability size did not moderate the effect of display type, $F(4,1902)=1.49$, $p=.20$, $\eta_p^2=.003$.[8]

Display type did not significantly affect understanding of relative risk magnitudes (foreground-only: *M*=0.89, *SD*=0.88; foreground+background: *M*=0.82, *SD*=0.88), *F*(1,1902)=2.26, *p*=.13, η_p^2=.001, and the non-significant trend was for foreground-only displays to produce greater understanding. There was no moderating effect of probability size, *F*(4,1902)=1.61, *p*=.17, η_p^2=.003.

3.2.3. RQ3: (a) What are the effects of display type on user evaluations, and (b) are any effects moderated by probability sizes?

Foreground-only displays (*M*=5.60, *SD*=1.19) were evaluated less positively than foreground+background displays (*M*=5.88, *SD*=1.06), *F*(1,1902)=30.70, *p*<.001, η_p^2=.02. This effect was not moderated by probability size, *F*(4,1902)=1.51, *p*=.20, η_p^2=.003.

3.3. Discussion

Experiment 2 replicated findings of Experiment 1 in a risk reduction scenario. Foreground-only effects for perceived risk and risk aversion were again weaker at larger probability sizes, and there were no effects on risk aversion for the two largest probability sizes. Experiment 2 also showed weaker foreground-only effects at larger probability sizes for participants' perceived decrease in risk due to the vaccines. Additionally, the magnitude of foreground-only effects was affected by the risk reduction level associated with each vaccine. When risk reduction was set to 20% (vs. 80%), participants' perceived decrease in risk was less affected by display type, and risk aversion was largely unaffected. These findings suggest that small risk reductions are seen as trivial independently of the display type.

Experiment 2 also revealed that foreground-only displays decreased understanding of absolute risk magnitudes, independently of probability size. These findings are in keeping with previous work showing that such displays hinder risk understanding (Garcia-Retamero & Galesic, 2010; Stone et al., 2015, 2017), and that this detrimental effect holds for understanding of absolute but not relative risk magnitudes (Stone et al., 2015).

Like Experiment 1, Experiment 2 showed that foreground-only displays received more negative user evaluations than foreground+background displays. A moderating effect of probability sizes was not observed in Experiment 2, suggesting that icon arrays may be liked less when they do not depict the background regardless of probability size. This finding contrasts with previous work that found no differences in user evaluations between foreground-only and foreground+background bar graphs (Bruine de Bruin et al., 2013; Okan et al., 2018; Stone et al., 2017). People may be relatively familiar with bar graphs that do not include background information, perhaps explaining why such graphs were not evaluated more negatively than foreground+background bar graphs in previous studies. Instead, people may expect icon arrays to include background information, leading to more negative evaluations when this is not depicted.

4. GENERAL DISCUSSION

Across two experiments, we provide the first empirical demonstration that the strength of foreground-only effects is contingent on depicted probability sizes and the risk reduction level associated with a protective behavior. Our findings show that the tendency for foreground-only displays to increase the perceived likelihood of a risky event and risk aversion (Chua et al., 2006; Hu et al., 2014; Okan et al., 2018; Schirillo & Stone, 2005; Stone et al., 1997, 2003, 2015, 2017) becomes less robust as the probability of the risky event increases. Foreground-only effects were also weaker in scenarios involving small relative risk reductions (i.e., where a vaccine only resulted in a small reduction in the risk of suffering a disease).

4.1. Implications for the design of graphical risk communications and decision support

Our results suggest that existing graph design recommendations should be adjusted for the depicted probability size. Although the effects on risk understanding and user evaluation were not typically moderated by probability size, the effects on perceived risk and risk aversion were. Thus, it is worth considering the recommendations for low-probability risks and large-probability risks separately.

For small probabilities, foreground-only displays can encourage risk aversion, but at the cost of reduced understanding of absolute risk magnitudes (see also Stone et al., 2015, 2017). Recent work, however, has shown that the detrimental effect of foreground-only bar graphs on understanding can be eliminated by adding simple numerical labels containing information on the foreground and background above the bars (Okan et al., 2018). It would be useful to test whether similar labels can eliminate the negative impact of foreground-only displays on understanding for other graph types (e.g., labels placed next to icon arrays; see e.g., Okan et al., 2015). If this is the case, foreground-only displays may be a viable risk-communication technique for low-probability risks if the goal is to promote risk aversion, albeit one that should be used very cautiously. If this is not the case, foreground+background displays should be used instead, as foreground-only displays could compromise informed decision making, making their use ethically questionable.

When probabilities are larger, foreground-only displays are less likely to encourage risk aversion relative to foreground+background displays, yet still hinder understanding of absolute risk magnitudes. Our results also showed that foreground-only icon arrays are evaluated more negatively. People may not be motivated to attend to, or take actions regarding, communications that they dislike (Ancker et al., 2006; Bruine de Bruin et al., 2013; Okan et al., 2018; Stone et al., 2017). Thus, for larger probabilities, the disadvantages of foreground-only displays remain (i.e., a negative impact on understanding and user evaluations), while the potential advantages are minimized or disappear entirely (i.e., encouraging risk aversion). Hence, our work suggests that foreground-only displays should be avoided entirely for high probability risks.

Finally, our data also suggest that foreground-only displays are less likely to increase risk aversion in scenarios involving only small risk reductions, regardless of probability size. Thus, in such cases foreground-only displays may not promote risk aversion even if probabilities are small.

4.2. Implications for theory and research

Existing theoretical accounts of foreground-only effects share the idea that foreground-only displays hinder people's ability to see that a risk is small when this is indeed the case (Ancker et al., 2006; Stone et al., 2003, 2018). Our findings support this idea while refining and extending these accounts to situations involving a wide range of probability sizes. Specifically, our results indicate that increases in probability size are more prone to result in increases in perceived risk for foreground+background displays than for foreground-only displays.

Our results also have implications for the conceptualization of low vs. high probabilities. Previous work has often highlighted the need to determine whether results documented with low-probability risks also hold for larger probabilities (Stone et al., 2017) without defining what was meant by low vs. large. Moreover, although some authors have defined low probabilities as less than 1% (Gurmankin, Helweg-Larsen, Armstrong, Kimmel, & Volpp, 2005; Lipkus, 2007), studies focusing on low probability risks have often used somewhat greater probabilities (Table I). Our results suggest that a binary conceptualization of “low” vs. “large” may be misguided. Instead, a graduated conceptualization may be more meaningful, because probability size shows a continuous relationship to our outcomes and the size of foreground-only effects (Figs. 2 & 4).

4.3. Limitations and future research

Our work has some limitations. First, although we tested probabilities that substantially exceeded those examined in previous studies, our largest probability was below 50%. Hence the number of affected individuals was always smaller than the number of unaffected ones. Future studies could examine foreground-only effects for probabilities of 50% or above to determine if the trends depicted in Figs. 2 and 4 continue to hold.

Second, to manipulate probability size we altered the number of affected individuals (i.e., the numerators), while keeping the number of people at risk (i.e., the denominators) constant and set to 1,000. Thus, it is not clear whether our findings would hold with smaller denominators. Although some studies on foreground-only effects used denominators as small as 50 (e.g., Bruine de Bruin et al., 2013) or 100 (Hu et al., 2014; Shepperd et al., 2013; Stone et al., 2017), they did not involve large probability sizes. Hence future research could examine foreground-only effects for different combinations of probability and denominator sizes. Based on existing theoretical accounts of foreground-only effects, we expect that the moderating effect of probability size should hold independently of denominator size. It seems likely, however, that risk understanding would be better for smaller denominators, as smaller group sizes (e.g., 100 to 200) can be easier to process and imagine than larger group sizes (Garcia-Retamero & Galesic, 2011).

Third, our icon arrays did not include groupings (e.g., of 100 units) that could have potentially facilitated counting (see e.g., Galesic et al., 2009). However, considering existing theoretical accounts we also have no reason to expect that this design feature would impact our finding that foreground-only effects decrease with larger probability sizes. That is, the salience of the foreground relative to the background and the availability of the part-to-whole relationship should not be affected by whether icon arrays include groupings or not.

Finally, in the current work we measured behavioral intentions in hypothetical medical scenarios. Future work should investigate whether our findings generalize to actual behavior in contexts relevant for people’s well-being, including medical, financial, or environmental settings.

4.4. Conclusions

Our findings indicate that the strength of a well-documented graphical display effect (the “foreground-only effect”) is contingent on two key factors: the depicted probability size and the risk reduction level associated with a protective behavior. As probability sizes increases, the tendency for foreground-only displays to produce greater perceived risk and risk aversion than foreground+background displays is reduced. Foreground-only effects also become weaker for smaller risk reductions. Our work emphasizes the importance of considering these two factors in theoretical and applied projects on graphical risk communication and decision support.

REFERENCES

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, *13*, 608–618.

Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint versus separate evaluation. *Journal of Economic Behavior & Organization*, *39*, 41–58.

Berry, J. D., Dyer, A., Cai, X., Garside, D. B., Ning, H., Thomas, A., … Lloyd-Jones, D. M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, *366*, 321–329.

Bruine de Bruin, W., Stone, E. R., Gibson, J. M., Fischbeck, P. S., & Shoraka, M. B. (2013). The effect of communication design and recipients' numeracy on responses to UXO risk. *Journal of Risk Research*, *16*, 981–1004.

Chua, H. F., Yates, J. F., & Shah, P. (2006). Risk avoidance: Graphs versus numbers. *Memory & Cognition*, *34*, 399–410.

Ferrer, R. A., Klein, W. M., Persoskie, A., Avishai-Yitshak, A., & Sheeran, P. (2016). The tripartite model of risk perception (TRIRISK): Distinguishing deliberative, affective, and experiential components of perceived risk. *Annals of Behavioral Medicine*, *50*, 653–663.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*, 1–17.

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

Garcia-Retamero, R., & Cokely, E. T. (2011). 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*, *17*, 270–287.

Garcia-Retamero, R., & Cokely, E. T. (2014). The influence of skills, message frame, and visual aids on prevention of sexually transmitted diseases. *Journal of Behavioral Decision Making*, *27*, 179–189.

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

Garcia-Retamero, R., & Galesic, M. (2011). Using plausible group sizes to communicate information about medical risks. *Patient Education and Counseling*, *84*, 245–250.

Gurmankin, A. D., Helweg-Larsen, M., Armstrong, K., Kimmel, S. E., & Volpp, K. G. M. (2005). Comparing the standard rating scale and the magnifier scale for assessing risk perceptions. *Medical Decision Making*, *25*, 560–570.

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

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

Kreuzmair, C., Siegrist, M., & Keller, C. (2017). Does iconicity in pictographs matter? The influence of iconicity and numeracy on information processing, decision making, and liking in an eye-tracking study. *Risk Analysis*, *37*, 546–556.

Larson, H. J., Jarrett, C., Schulz, W. S., Chaudhuri, M., Zhou, Y., Dube, E., … Wilson, R. (2015). Measuring vaccine hesitancy: The development of a survey tool. *Vaccine*, *33*, 4165–4175.

Lipkus, I. M. (2007). Numeric, verbal, and visual formats of convening health risks: Suggested best practices and future recommendations. *Medical Decision Making*, *27*, 695–713.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*, 390–401.

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

Okan, Y., Stone, E. R., & Bruine de Bruin, W. (2018). Designing graphs that promote both risk understanding and behavior change. *Risk Analysis*, *38*, 929–946.

Okan, Y., Stone, E. R., Parillo, J., Bruine de Bruin, W., & Parker, A. (2019, November 4). Probability size matters: The effect of foreground-only vs. foreground+background graphs on risk aversion diminishes with larger probabilities. Dataset. https://doi.org/10.17605/OSF.IO/UYRWF

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

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89–107.

Schirillo, J. A., & Stone, E. R. (2005). The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Analysis*, *25*, 555–566.

Schmiege, S. J., Bryan, A., & Klein, W. M. (2009). Distinctions between worry and perceived risk in the context of the theory of planned behavior. *Journal of Applied Social Psychology*, *39*, 95–119.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*, 966–972.

Shepperd, J. A., Lipkus, I. M., Sanderson, S. C., McBride, C. M., O’Neill, S. C., & Docherty, S. (2013). Testing different communication formats on responses to imagined risk of having versus missing the GSTM1 gene. *Journal of Health Communication*, *18*, 124–137.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*, 311–322.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1978). Accident probabilities and seat belt usage: A psychological perspective. *Accident Analysis and Prevention*, *10*, 281–285.

Spiegelhalter, D. (2017). Risk and uncertainty communication. *Annual Review of Statistics and Its Application*, *4*, 31–60.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*, 1393–1400.

Stone, E. R., Bruine de Bruin, W., Wilkins, A. M., Boker, E. M., & MacDonald Gibson, J. (2017). Designing graphs to communicate risks: Understanding how the choice of graphical format influences decision making. *Risk Analysis*, *37*, 612–628.

Stone, E. R., Gabard, A. R., Groves, A. E., & Lipkus, I. M. (2015). Effects of numerical versus foreground-only icon displays on understanding of risk magnitudes. *Journal of Health Communication*, *20*, 1230–1241.

Stone, E. R., Reeder, E. C., Parillo, J., Long, C., & Walb, L. (2018). Salience versus proportional reasoning: Rethinking the mechanism behind graphical display effects. *Journal of Behavioral Decision Making*, *31*, 473–486.

Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, *90*, 19–36.

Stone, E. R., Yates, J. F., & Parker, A. M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, *3*, 243–256.

Trevena, L. J., Zikmund-Fisher, B. J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P. K., … Woloshin, S. (2013). Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. *BMC Medical Informatics and Decision Making*, *13*, S7.

Zikmund-Fisher, B. J., Ubel, P., Smith, D., Derry, H., McClure, J., Stark, A., … Fagerlin, A. (2008). Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. *Patient Education and Counseling*, *73*, 209–214.

Zikmund-Fisher, B. J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Kahn, V. C., Exe, N. L., … Fagerlin, A. (2014). Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, *34*, 443–453.

Zipkin, D. A., Umscheid, C. A., Keating, N. L., Allen, E., Aung, K., Beyth, R., … Feldstein, D. A. (2014). Evidence-based risk communication: A systematic review. *Annals of Internal Medicine*, *161*, 270–280.

Footnotes

[1] Of note, studies generally assume a continuum between risk aversion and risk taking. Foreground-only displays tend to drive people towards risk aversion, which is often treated as desirable since it is deemed good to be safe. However, in practice this can be highly context-dependent and should consider recipients' personal values as well as any other potential effects of an action (e.g., side effects of the risk-mitigating action).

[2] Perceived risk can include both cognitive and affective components such as worry (Finucane, Alhakami, Slovic, & Johnson, 2000; Slovic, Finucane, Peters, & MacGregor, 2004). These two components can be differentially related to key outcomes such as behavioral intentions (Ferrer, Klein, Persoskie, Avishai-Yitshak, & Sheeran, 2016; Schmiege, Bryan, & Klein, 2009). However, in the current study cognitive risk perception and worry were highly correlated and were similarly affected by our experimental manipulations. Thus, our composite measures of perceived risk and perceived decrease in risk include both of these factors. Supplementary materials include separate results for cognitive risk perceptions and worry (Tables S1-S3).

[3] The first user evaluation item (perceived understanding) was presented separately from the other user evaluation items because we considered that it could potentially also be associated with objective risk understanding (e.g., Stone et al., 2015). Analyses revealed that the item was highly correlated with the other user evaluation items (*r*s=.68, .50, and .60) and only weakly correlated with risk understanding (r=.15). Hence, we averaged this item with the remaining user evaluation items (see Bruine de Bruin et al., 2013; Okan et al., 2018, for a similar approach).

[4] In Experiment 1, inattentive participants were excluded based on the total time the survey was open. In Experiment 2, we aimed to achieve an improved measure of inattention by considering only the time that participants spent on the survey itself, not including the consent form, comments page (which was only displayed in Experiment 2), and MTurk code entry.

[5] Participants were also asked to assess the chances of contracting Slibitis and how worried they would be about contracting Slibitis *with* each of the vaccines. These items were based on previous work (Shepperd et al., 2013) and were included in the current study for exploratory purposes. Our exploratory analyses appear in the Supplementary materials, and suggested that results with this measure are ambiguous.

[6] Participants' estimates of the relative risk reduction achieved by each vaccine were computed by subtracting the estimated number of people affected with the vaccine from the estimated number of people affected without the vaccine, and dividing the resulting value by the estimated number of people affected without the vaccine. This procedure was applied separately for each vaccine. Estimates of the risk reduction achieved by one vaccine relative to another were computed by subtracting the number of people affected with the less effective vaccine from the more effective vaccine, and dividing the resulting value by the estimated number of people affected with the more effective vaccine. Because getting two estimates correct mathematically guaranteed that the third would be correct, we scored understanding of the relative risk magnitudes as 0=none correct, 1=one correct, and 2=all correct.

[7] Main effects of order were observed in all cases except for analyses of user evaluations and understanding of absolute risk magnitudes, indicating that participants presented with the 20% risk reduction first were more risk averse than

those presented with the 80% risk reduction first. This main effect is to be expected, given that one's initial judgment of the risk may change considerably after viewing another display with a different risk reduction (Bazerman, Moore, Tenbrunsel, Wade-Benzoni, & Blount, 1999). For example, if a participant views the 20% risk reduction first, they may substantially increase their assessment of the risk after seeing that the other vaccine triples the reduction in risk. Of the 16 potential interactions involving order and display type for all the dependent variables, only one was significant, which is consistent with chance. We therefore do not discuss the order variable further. All effects involving order can be found in Supplementary materials (Table S7).

[8] In an additional analysis where only the exact answers were treated as correct (vs. answers within 25% of the correct value), the main effect of display type did not reach significance. Additionally, there was a small (but significant) interaction between probability size and display type, with foreground-only displays producing worse understanding only for smaller probabilities (Table S5, Supplementary materials). This can be explained considering that our displays did not provide numerical information about the number of people affected vs. not affected. Hence, getting the exact correct response is particularly difficult in conditions involving probability sizes of 10% and above, where performance was poor for both display types. An analysis treating answers as correct if within 50% of the correct value (to better capture participants' understanding of the approximate magnitude, i.e., the gist) yielded results in line with our original scoring criterion, with a stronger main effect of display type and no interaction.

Table I. Probability sizes and risk reduction used in previous studies investigating effects of foreground-only vs. foreground+background displays on perceived risk, perceived decrease in risk, and risk aversion.

Study	**Probability sizes**	**Relative risk reduction**	**Outcome measure**	**Effect size (Cohen's d)**
Bruine de Bruin et al. (2013)	7.6% vs. 4.8%	37%	Risk aversion (support to increase worker safety)	.26
			Risk taking (support for construction at risky site)	.20
Hu et al. (2014) [a]	7 % vs. 3.5%	50%	Risk aversion (preference for driving on safer road)	Study 1: .39 Study 3: .59
Okan et al. (2018) [b]	8% vs. 4%	50%	Risk aversion (willingness to take a drug)	.19
			Perceived risk (perceived risk of suffering heart attack & of reduction in risk achieved by treatment)	.59
Shepperd et al. (2013)	11% vs. 9%	18%	Perceived risk (estimated risk of getting lung cancer *with* a genetic marker)	-.76
			Perceived risk (estimated risk of getting lung cancer *without* a genetic marker)	-. 19

Stone et al. (2003)	.6% vs. .3%	50%	Risk aversion (willingness to pay for safer toothpaste)	Study 1: .65 Study 2: .52
			Perceived decrease in risk (perceived risk reduction size & significance)	Study 1: .90 Study 2: .88
Stone et al. (2017)	0% - 10%	n/a	Risk aversion (support to increase worker safety)	.02
			Risk taking (support for construction at risky site)	-.19
			Perceived risk (estimated risk of being exposed to unexploded ammunition)	.31

Note: Where more than one item was used to assess the outcome measures, we averaged across all relevant outcomes for each measure to compute Cohen's d. In studies that manipulated additional factors besides foreground only vs. foreground+background display type, we collapsed across levels of the additional factors, if interactions with display type did not exist. When interactions existed, we focused on the condition that was most comparable to previous studies on foreground-only effects. [a] Results are reported for the "default processing mindset" condition in Study 3; [b] Results are reported for the condition that did not include additional numerical labels, for perceived risk.

Table II. Experiment 1 mean risk aversion and perceived risk for foreground-only and foreground+background displays as a function of probability size.

Probability size	F.O.	F.B.	Mean diff	Cohen's d	*t*	*p*
			Risk Aversion			
.1%	-.04 (.81)	-.40 (.79)	.36	0.45	4.27	< .001
1%	-.06 (.73)	-.27 (.69)	.21	0.29	2.71	.007
10%	.07 (.71)	.04 (.70)	.03	0.05	0.45	.65
20%	.17 (.73)	.07 (.72)	.10	0.13	1.27	.20
40%	.13 (.74)	.26 (.71)	-.13	-0.17	-1.65	.10
			Perceived Risk			
.1%	3.36 (1.45)	2.03 (1.16)	1.33	1.01	9.48	< .001
1%	3.00 (1.54)	2.08 (0.86)	0.92	0.74	7.04	< .001
10%	3.86 (1.59)	2.99 (1.16)	0.87	0.63	5.97	< .001
20%	4.31 (1.57)	3.22 (1.11)	1.09	0.80	7.63	< .001
40%	4.27 (1.53)	3.94 (1.22)	0.33	0.23	2.24	.03

Note. F.O. = Foreground-only mean, F.B. = Foreground+background mean.

Standard deviations in parentheses. Mean risk aversion scores are in z-score units.

Table III. Number of people out of 1,000 affected by Slibitis for all probability size and risk reduction combinations in Experiment 2.

	Number of people out of 1,000 affected by Slibitis		
Probability size	Without vaccine	20% risk reduction vaccine	80% risk reduction vaccine
.5%	5	4	1
1%	10	8	2
10%	100	80	20
20%	200	160	40
40%	400	320	80

Table IV. Experiment 2 mean risk aversion, perceived risk, and perceived decrease in risk for foreground-only and foreground+background displays as a function of probability size and risk reduction level

	20% Risk Reduction						80% Risk Reduction					
Probability size	F.O.	F.B.	Mean diff.	Cohen's d	*t*	*p*	F.O.	F.B.	Mean diff.	Cohen's d	*t*	*p*
	Risk Aversion											
.5%	2.50 (1.41)	2.41 (1.42)	0.09	0.07	0.46	.65	4.22 (1.54)	3.77 (1.61)	0.45	0.29	2.86	.005
1%	2.80 (1.51)	2.45 (1.40)	0.35	0.24	2.18	.03	4.61 (1.41)	3.88 (1.41)	0.73	0.52	5.36	< .001
10%	3.29 (1.75)	3.03 (1.55)	0.26	0.16	1.73	.09	4.92 (1.34)	4.41 (1.53)	0.51	0.36	4.36	< .001
20%	3.24 (1.62)	3.24 (1.58)	0.00	0.00	0.55	.58	4.79 (1.42)	4.69 (1.37)	0.10	0.07	0.22	.82
40%	3.12 (1.67)	3.21 (1.69)	-0.09	-0.06	0.16	.87	4.90 (1.30)	4.95 (1.37)	-0.05	-0.04	0.08	.94
	Perceived Risk											
.5%	3.41 (1.86)	2.41 (1.43)	1.00	0.60	6.17	< .001	3.59 (1.76)	2.60 (1.43)	0.99	0.62	6.52	< .001

1%	3.58 (1.75)	2.56 (1.24)	1.02	0.67	6.26	< .001	3.79 (1.75)	2.67 (1.23)	1.12	0.73	6.94	< .001
10%	4.28 (1.57)	3.41 (1.43)	0.87	0.58	6.79	< .001	4.52 (1.52)	3.60 (1.45)	0.92	0.62	7.00	< .001
20%	4.43 (1.51)	3.76 (1.43)	0.67	0.46	4.50	< .001	4.57 (1.54)	3.81 (1.44)	0.76	0.51	4.99	< .001
40%	4.77 (1.44)	4.18 (1.37)	0.59	0.43	4.78	< .001	4.95 (1.40)	4.49 (1.33)	0.46	0.34	4.16	< .001
	Perceived Decrease in Risk											
.5%	2.37 (1.21)	2.11 (1.24)	0.26	0.21	2.04	.04	4.45 (1.71)	3.74 (1.80)	0.71	0.40	4.11	< .001
1%	2.42 (1.29)	2.11 (1.12)	0.31	0.26	2.17	.03	4.76 (1.70)	3.84 (1.58)	0.92	0.56	5.49	< .001
10%	3.03 (1.34)	2.79 (1.37)	0.24	0.18	2.25	.03	5.07 (1.37)	4.55 (1.59)	0.52	0.35	3.98	< .001
20%	2.90 (1.37)	2.90 (1.44)	0.00	0.00	.29	.78	5.04 (1.44)	4.61 (1.45)	0.43	0.30	2.57	0.009
40%	2.86 (1.31)	2.98 (1.35)	-0.12	-0.09	.78	.43	5.07 (1.27)	4.96 (1.37)	0.11	0.09	.89	.37

Note. F.O. = Foreground-only mean, F.B. = Foreground+background mean. Standard deviations in parentheses.

Fig. 1. Icon arrays in Experiment 1 in the foreground-only (a & c) and foreground+background (b & d) conditions, for probability size = 1% (a & b) and probability size = 40% (c & d).

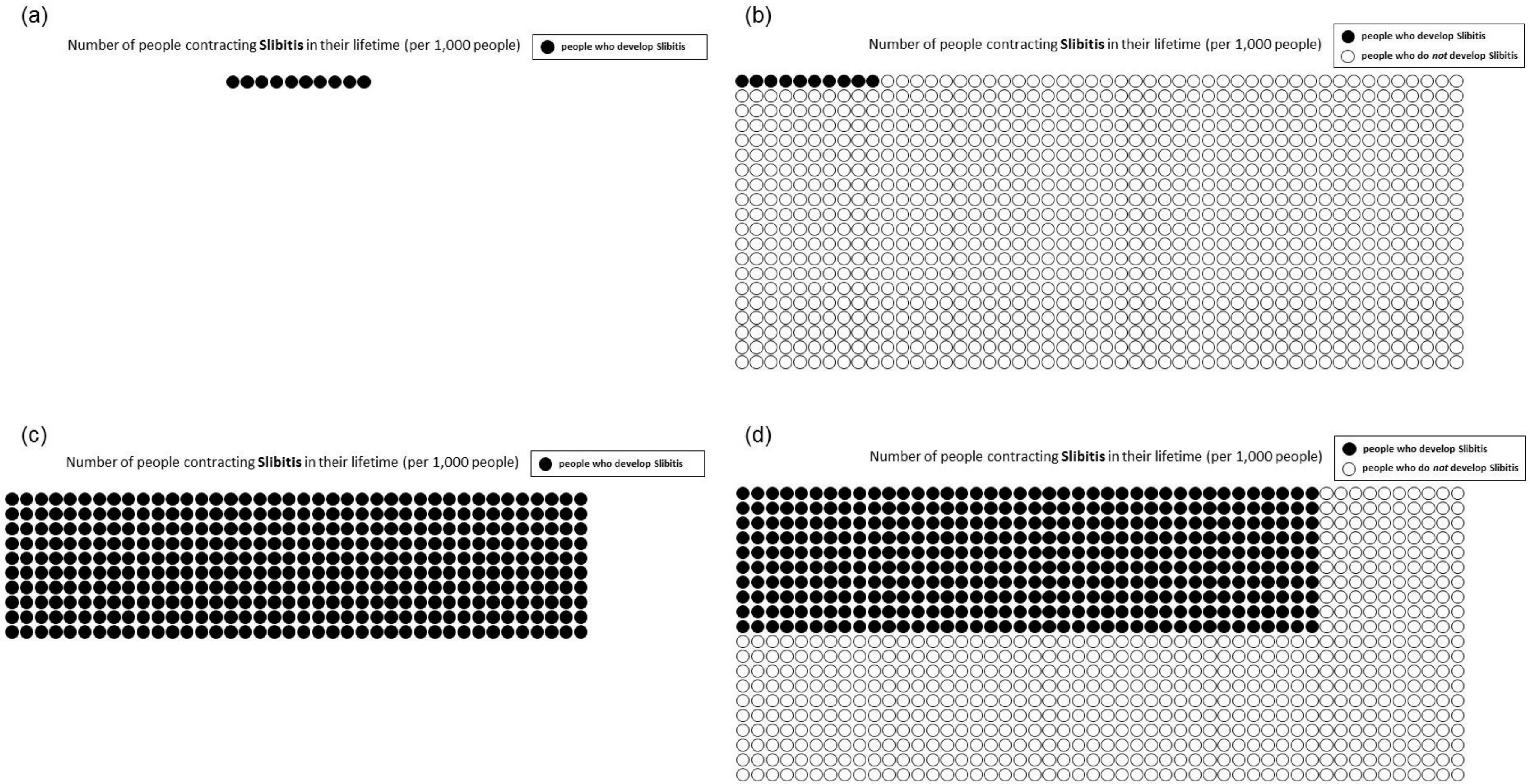
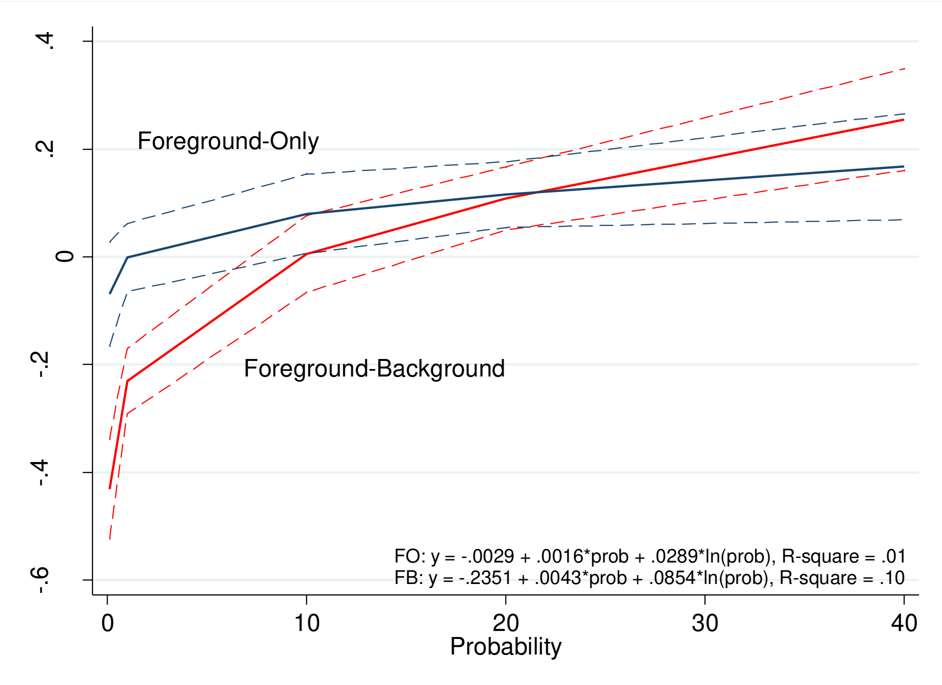


Fig. 2. Experiment 1 mean predicted (a) risk aversion and (b) perceived risk with 95% confidence intervals. Note: Solid line depicts mean predicted values, estimated using linear regression with probability size and its natural logarithm as predictors, with dashed lines representing the 95% confidence interval.

(a) Risk Aversion



(b) Perceived Risk

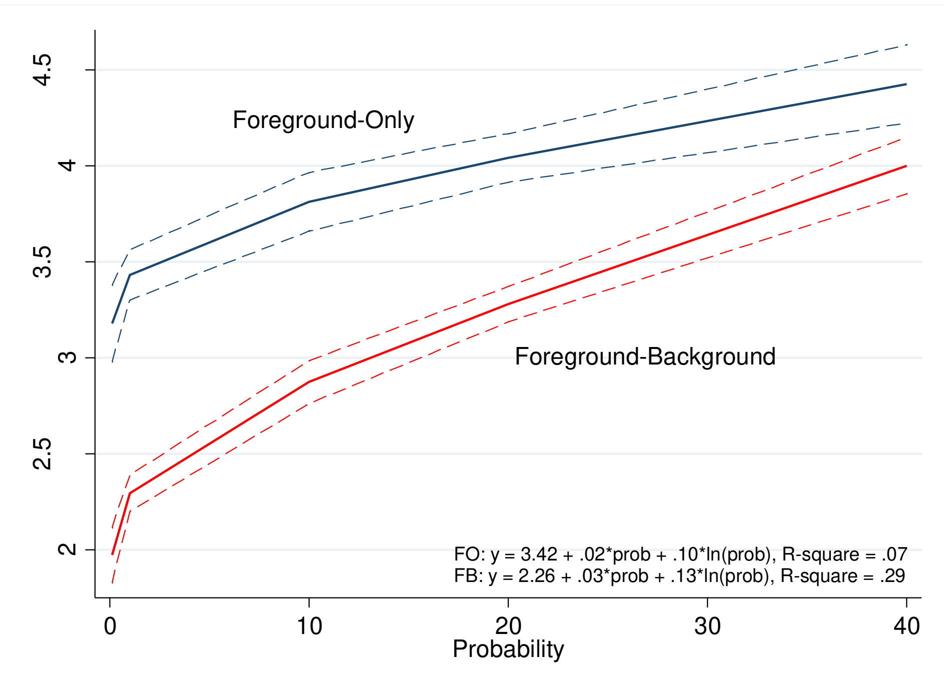


Fig. 3. Icon arrays in Experiment 2 in the (a) foreground-only and (b) foreground+background conditions, for probability size = 1% and risk reduction = 80%

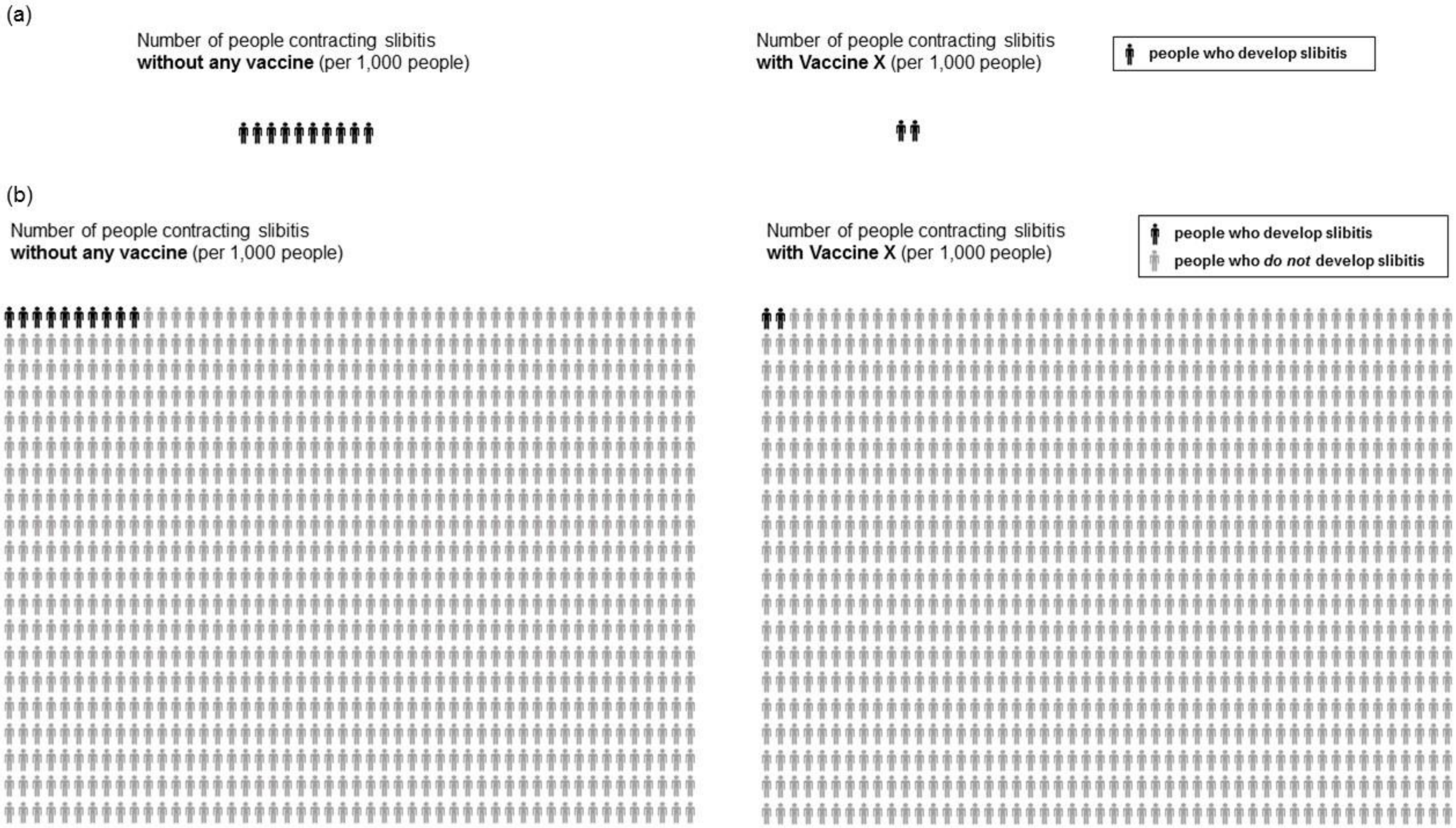
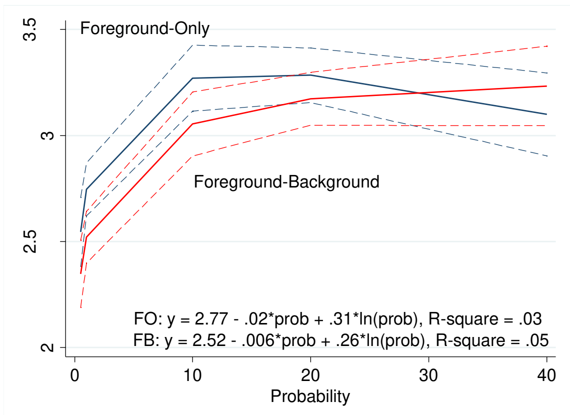


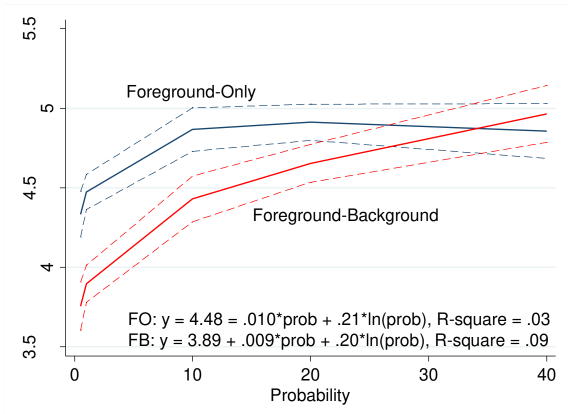
Fig. 4. Experiment 2 mean predicted (a) risk aversion, (b) perceived risk, (c) perceived decrease in risk, with 95% confidence intervals. Note: Solid line depicts mean predicted values, estimated using linear regression with probability size and its natural logarithm as predictors, with dashed lines representing the 95% confidence interval.

20% Risk Reduction

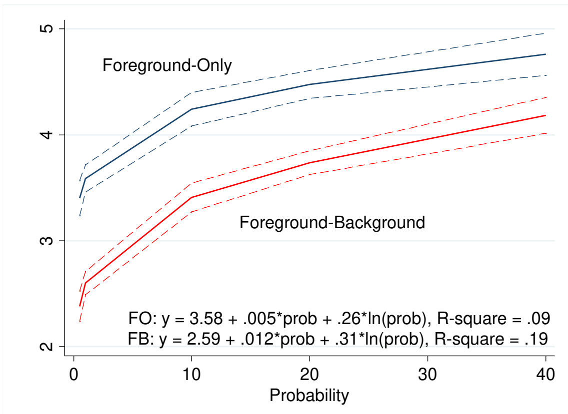
80% Risk Reduction

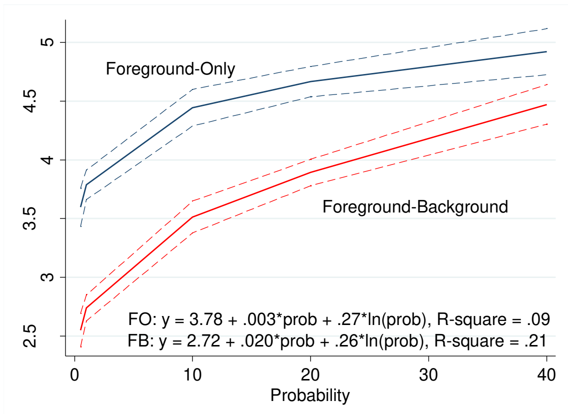
(a) Risk Aversion



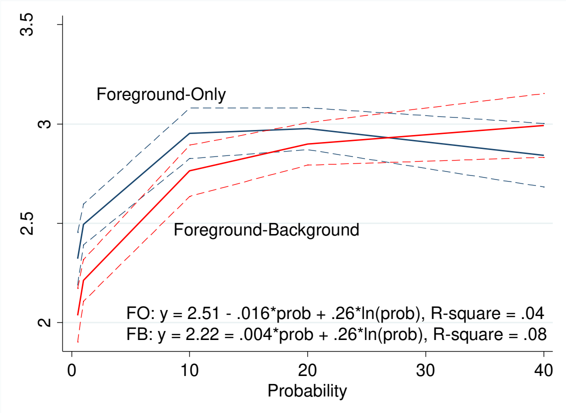


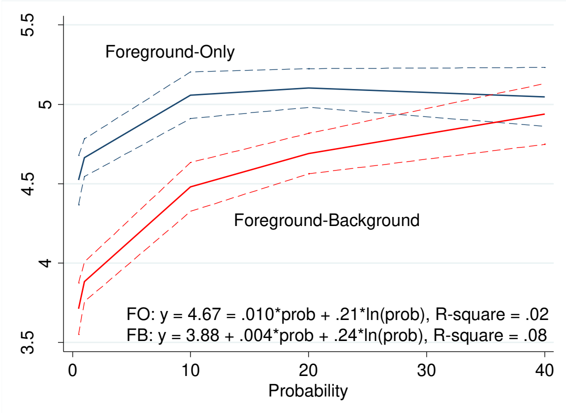
(b) Perceived Risk





(c) Perceived Decrease in Risk





Supplementary materials for Risk Analysis

Probability size matters: The effect of foreground-only vs. foreground+background graphs on risk aversion diminishes with larger probabilities

Yasmina Okan,[1] * Eric R. Stone,[2] Jonathan Parillo,[3] Wändi Bruine de Bruin,[1,4] Andrew M. Parker [5]

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

[2]Department of Psychology, Wake Forest University, Winston-Salem, NC, USA

[3]Department of Psychology, Vanderbilt University, Nashville, TN, USA

[4]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA

[5]RAND Corporation, Pittsburgh, PA, USA

* Address correspondence to Yasmina Okan, Centre for Decision Research, Leeds University Business School, University of Leeds, Charles Thackrah Building, Leeds, LS2 9LB, UK, Phone: +44 113 343 2622; E-mail: y.okan@leeds.ac.uk

Results for individual items corresponding to the outcome measures assessed

Table S1. Mean scores in Experiment 1, as a function of probability sizes and display type (SD in parentheses).

	.1%		1%		10%		20%		40%	
Item	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.
Risk aversion										
1. Recommendation	5.13 (1.70)	4.56 (1.81)	5.15 (1.65)	4.92 (1.63)	5.42 (1.47)	5.35 (1.60)	5.55 (1.53)	5.43 (1.65)	5.46 (1.45)	5.56 (1.59)
2. Budget CDC	31.25 (22.23)	23.48 (19.97)	30.99 (20.82)	24.29 (19.30)	32.85 (22.21)	33.17 (21.17)	35.91 (22.41)	31.92 (20.88)	34.63 (21.98)	38.71 (20.33)
3. Education CDC	4.37 (1.36)	3.90 (1.35)	4.26 (1.16)	4.07 (1.21)	4.46 (1.19)	4.38 (1.21)	4.56 (1.16)	4.51 (1.37)	4.55 (1.27)	4.70 (1.14)
Perceived risk										
1.Chances	3.45 (1.44)	1.95 (1.14)	3.05 (1.58)	2.02 (.88)	4.05 (1.72)	2.93 (1.10)	4.42 (1.61)	3.14 (1.03)	4.44 (1.67)	3.92 (1.13)
2. Worry	3.28 (1.64)	2.12 (1.32)	2.96 (1.63)	2.14 (1.09)	3.67 (1.71)	3.05 (1.44)	4.20 (1.73)	3.30 (1.43)	4.09 (1.70)	3.96 (1.55)
Risk understanding										
1.People expected to contract Slibitis	.30 (46)	.64 (.48)	.54 (.50)	.47 (.50)	.50 (.50)	.31 (.47)	.51 (.50)	.34 (.48)	.49 (.50)	.53 (.50)
User evaluations										
1.Perceived understanding	3.52 (2.22)	6.35 (1.04)	5.63 (1.60)	6.28 (.94)	5.03 (1.87)	6.22 (1.04)	5.08 (1.97)	5.92 (1.21)	4.81 (1.91)	5.72 (1.22)
2. Perceived helpfulness	2.83 (1.86)	5.51 (1.59)	4.40 (1.87)	5.55 (1.36)	3.98 (2.13)	5.41 (1.49)	4.15 (1.98)	5.08 (1.64)	3.61 (1.90)	5.06 (1.73)

3. Liking	2.60 (1.75)	5.05 (1.78)	3.41 (1.95)	4.69 (1.86)	3.01 (2.06)	4.36 (2.04)	2.83 (1.88)	3.95 (1.95)	2.36 (1.69)	3.96 (2.03)
4. Trust	2.93 (1.79)	5.24 (1.48)	4.25 (1.75)	5.13 (1.42)	3.68 (1.87)	4.94 (1.46)	3.76 (1.71)	4.76 (1.51)	3.34 (1.63)	4.51 (1.59)

Table S2. Mean scores in Experiment 2 for the 20% Risk Reduction condition, as a function of probability sizes and display type (SD in parentheses).

	.5%		1%		10%		20%		40%	
Item	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.
Risk aversion										
1. Recommendation	2.88 (1.81)	2.74 (1.69)	3.22 (1.86)	2.90 (1.78)	3.81 (2.05)	3.66 (1.94)	3.86 (1.94)	3.85 (1.91)	3.62 (1.99)	3.66 (1.97)
2. Wait for vaccination	2.63 (1.81)	2.60 (1.83)	3.07 (1.94)	2.63 (1.82)	3.45 (2.16)	3.16 (1.95)	3.45 (1.99)	3.44 (2.04)	3.36 (2.02)	3.52 (2.21)
3. Willingness to pay	1.99 (1.18)	1.88 (1.19)	2.12 (1.21)	1.82 (1.08)	2.60 (1.58)	2.26 (1.23)	2.40 (1.41)	2.43 (1.3)	2.37 (1.44)	2.45 (1.44)
Perceived risk										
1. Chances	3.46 (1.93)	2.30 (1.48)	3.71 (1.89)	2.54 (1.32)	4.44 (1.60)	3.37 (1.40)	4.63 (1.52)	3.76 (1.40)	5.02 (1.46)	4.12 (1.37)
2. Worry	3.35 (1.95)	2.53 (1.64)	3.45 (1.88)	2.58 (1.43)	4.12 (1.81)	3.44 (1.71)	4.23 (1.71)	3.75 (1.65)	4.53 (1.69)	4.23 (1.6)
Perceived decrease in risk										
1. Chances	2.38 (1.22)	2.16 (1.24)	2.52 (1.30)	2.24 (1.12)	3.05 (1.42)	2.79 (1.34)	2.91 (1.34)	2.92 (1.43)	2.96 (1.44)	2.92 (1.37)
2. Worry	2.35 (1.58)	2.06 (1.51)	2.33 (1.57)	1.98 (1.43)	3.02 (1.7)	2.80 (1.71)	2.89 (1.65)	2.89 (1.76)	2.76 (1.55)	3.05 (1.69)
Perceived risk with vaccine										
1. Chances	3.19 (1.75)	2.16 (1.31)	3.24 (1.68)	2.47 (1.41)	3.61 (1.59)	2.87 (1.32)	3.69 (1.38)	3.17 (1.35)	4.10 (1.51)	3.54 (1.32)
2. Worry	3.21 (1.78)	2.38 (1.52)	3.18 (1.68)	2.57 (1.47)	3.59 (1.63)	3.03 (1.56)	3.62 (1.6)	3.22 (1.47)	3.89 (1.63)	3.67 (1.43)

Table S3. Mean scores in Experiment 2 for the 80% Risk Reduction condition, as a function of probability sizes and display type (SD in parentheses)

	.5%		1%		10%		20%		40%	
Item	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.
Risk aversion										
1. Recommendation	5.19 (1.67)	4.60 (1.88)	5.59 (1.55)	4.76 (1.61)	5.89 (1.39)	5.47 (1.69)	5.74 (1.56)	5.72 (1.48)	5.94 (1.21)	5.86 (1.41)
2. Wait for vaccination	4.31 (2.05)	3.88 (2.05)	4.79 (1.91)	4.06 (1.97)	5.12 (1.82)	4.51 (2.04)	5.08 (1.83)	4.87 (1.85)	5.17 (1.8)	5.26 (1.8)
3. Willingness to pay	3.17 (1.48)	2.84 (1.41)	3.44 (1.39)	2.82 (1.27)	3.75 (1.51)	3.24 (1.44)	3.53 (1.41)	3.48 (1.41)	3.60 (1.43)	3.74 (1.48)
Perceived risk										
1. Chances	3.68 (1.86)	2.54 (1.48)	3.87 (1.86)	2.55 (1.19)	4.54 (1.56)	3.55 (1.46)	4.70 (1.56)	3.79 (1.44)	5.09 (1.43)	4.48 (1.36)
2. Worry	3.50 (1.91)	2.65 (1.59)	3.70 (1.83)	2.80 (1.47)	4.50 (1.71)	3.64 (1.68)	4.43 (1.7)	3.84 (1.65)	4.81 (1.69)	4.50 (1.54)
Perceived decrease in risk										
1. Chances	4.58 (1.75)	3.73 (1.81)	4.79 (1.69)	3.94 (1.62)	5.17 (1.37)	4.63 (1.63)	5.10 (1.47)	4.68 (1.50)	5.26 (1.30)	5.08 (1.43)
2. Worry	4.31 (2.09)	3.75 (2.14)	4.73 (1.96)	3.74 (1.96)	4.98 (1.74)	4.47 (2.03)	4.98 (1.77)	4.54 (1.86)	4.89 (1.74)	4.84 (1.74)
Perceived risk with vaccine										
1. Chances	1.71 (1.12)	1.40 (0.94)	1.62 (1.08)	1.52 (0.99)	2.07 (1.1)	1.76 (1.13)	2.08 (1.08)	1.95 (1.15)	2.20 (1.02)	2.02 (0.97)
2. Worry	1.99 (1.4)	1.74 (1.21)	1.87 (1.19)	1.84 (1.11)	2.25 (1.22)	1.99 (1.11)	2.35 (1.3)	2.32 (1.43)	2.40 (1.28)	2.36 (1.2)

Table S4. Mean scores in Experiment 2 for understanding and user evaluation measures, as a function of probability sizes and display type (SD in parentheses)

	.5%		1%		10%		20%		40%	
	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.	F.O.	F.B.
Risk understanding										
1. Risk understanding—absolute	2.33 (1.20)	2.64 (0.86)	1.97 (1.34)	2.36 (1.03)	1.40 (1.27)	1.41 (1.21)	1.08 (1.21)	1.41 (1.22)	1.10 (1.23)	1.31 (1.15)
2. Risk understanding—relative	1.55 (0.79)	1.62 (0.70)	1.21 (0.89)	1.23 (0.9)	0.67 (0.78)	0.56 (0.73)	0.52 (0.69)	0.47 (0.67)	0.47 (0.66)	0.27 (0.55)
User evaluations										
1. Perceived understanding	6.40 (1.07)	6.61 (0.77)	6.30 (1.05)	6.53 (0.87)	6.23 (1.17)	6.24 (1.24)	6.14 (1.15)	6.28 (1.07)	6.19 (1.10)	6.31 (0.97)
2. Perceived helpfulness	6.15 (1.25)	6.48 (0.84)	6.13 (1.17)	6.29 (1.10)	5.94 (1.34)	6.03 (1.43)	5.75 (1.65)	6.07 (1.29)	5.94 (1.40)	6.11 (1.12)
3. Liking	5.48 (1.42)	6.09 (1.13)	5.03 (1.86)	5.62 (1.49)	4.88 (1.83)	5.28 (1.71)	4.95 (1.89)	5.46 (1.69)	5.12 (1.87)	5.29 (1.74)
4. Trust	5.13 (1.44)	5.75 (1.19)	5.14 (1.47)	5.45 (1.36)	4.96 (1.55)	5.16 (1.52)	4.96 (1.63)	5.31 (1.48)	5.12 (1.61)	5.23 (1.46)

Results for understanding in Experiment 2 using different scoring methods

Table S5. Mean understanding of absolute risk magnitudes with 3 different scoring methods, as a function of probability size and display type

Probability size	F.O.	F.B.	Mean diff.	Cohen's d	*t*	*p*
Correct = Exactly Correct						
.5%	2.22	2.45	-0.23	-0.20	-2.30	.02
1%	1.74	1.98	-0.24	-0.19	-1.73	.08
10%	1.03	0.85	0.18	0.16	1.25	.21
20%	0.54	0.63	-0.09	-0.10	-1.06	.29
40%	0.42	0.37	0.05	0.06	0.21	.84
Correct = Within 25%						
.5%	2.33	2.64	-0.31	-0.29	-3.14	.002
1%	1.97	2.36	-0.39	-0.32	-3.12	.002
10%	1.40	1.41	-0.01	-0.01	-0.46	.65
20%	1.08	1.41	-0.33	-0.27	-3.10	.002
40%	1.10	1.31	-0.21	-0.18	-1.89	.06
Correct = Within 50%						
.5%	2.34	2.66	-0.32	-0.31	-3.34	.001
1%	2.06	2.51	-0.45	-0.38	-3.73	.0002
10%	1.79	1.96	-0.17	-0.14	-1.67	.10
20%	1.59	1.88	-0.29	-0.24	-2.88	.004
40%	1.45	1.91	-0.46	-0.38	-4.11	.00005

In the analyses reported in the manuscript, responses were scored as correct if within 25% of the correct value. As reported in the manuscript, using this criterion there was a main effect of display type, $F(1,1902) = 23.27$, $p < .001$, $\eta_p^2 = .01$, and the interaction between display type and probability size was not significant, $F(4,1902) = 1.49$, $p = .20$, $\eta_p^2 = .003$.

When only the exact answer was treated as correct, the main effect of display type did not reach significance, $F(1, 1902) = 2.44$, $p = .12$, $\eta_p^2 = .001$, and there was a small (but significant) interaction with probability size, F (4, 1902) = 2.55, p = .04, η_p^2 = .005. Specifically, display type did not produce significant differences in understanding for probabilities of 10% or above, where performance was particularly poor (Table S5). The poorer performance at higher probabilities can be explained considering that our displays did not provide numerical information about the number of people affected vs. not affected (see Figure 3 in the manuscript). Hence, getting the exact correct response is particularly difficult in conditions involving probability sizes of 10% and above (regardless of display type), as this would require counting a large number of icons. Our original scoring criterion allows avoiding this issue by capturing participants' gist understanding. Indeed, when responses were coded as correct if within 50% of the correct value (to better capture gist understanding) there was a main effect of display type, F(1,1902) = 42.65, p < .001, η_p^2 = .02, and the interaction between display type and probability size was not significant, F(4,1902) = 0.92, p = .45, η_p^2 = .002, in line with the original scoring criterion.

One other striking feature of this analysis is that, regardless of probability size, the effect of display type gets considerably stronger as more gist-based measures of understanding are used. When participants were required to get the number exactly correct, there was only a .06 mean difference in understanding between the conditions ($M_{F\text{-}O}$ = 1.19 vs. $M_{F\text{-}B}$ = 1.25). When responses were coded as correct when within 25% of the correct value, the mean difference increased to .24 ($M_{F\text{-}O}$ = 1.58 vs. $M_{F\text{-}B}$ = 1.82). And when responses were coded as correct when within 50% of the correct value, the mean difference increased to .33 ($M_{F\text{-}O}$ = 1.85 vs. $M_{F\text{-}B}$ = 2.18). This finding suggests that foreground-only displays (compared to foreground+background displays) are particularly poor at capturing the gist of the absolute risk magnitudes.

Full ANOVA findings, Experiments 1 and 2

Table S6. Significant main effects or interactions not reported in main text, Exp. 1

Dependent Variable	Analysis	*F*	*p*
Risk Aversion	Main effect of Probability Size	21.60	<.001
Perceived Risk	Main effect of Probability Size	90.98	<.001
Understanding	Main effect of Probability Size	3.22	.012
User Evaluations	Main effect of Probability Size	15.09	<.001

Table S7. Significant main effects or interactions not reported in main text, Exp. 2

Dependent Variable	Analysis	*F*	*p*
Risk Aversion	Main effect of Probability Size	34.47	<.001
	Main effect of Risk Reduction	32.51	<.001
	Main effect of Order	161.63	<.001
	Risk Reduction x Probability Size	3.13	.014
	Risk Reduction x Order	192.77	<.001
	Risk Reduction x Probability Size x Order	3.32	.010
Perceived Risk	Main effect of Probability Size	88.38	<.001
	Main effect of Order	24.96	<.001
	Risk Reduction x Order	242.04	<.001
Perceived Decrease in Risk	Main effect of Probability Size	40.11	<.001
	Main effect of Risk Reduction	31.35	<.001
	Main effect of Order	158.09	<.001
	Risk Reduction x Order	10.00	.002
Risk understanding—absolute	Main effect of Probability Size	94.72	<.001
Risk understanding—relative	Main effect of Probability Size	193.19	<.001
	Main effect of Order	8.74	.003
	Display Type x Probability Size x Order	5.42	<.001
User Evaluations	Main effect of Probability Size	11.17	<.001

Additional exploratory analyses for items assessing perceived risk *with* each of the vaccines

Items assessing perceived risk *with* the vaccines were included for exploratory purposes, based on previous work (e.g., Shepperd et al., 2013). We reasoned that high values on this measure can be interpreted in two different ways: (1) as evidence of *high* perceived risk (and thus indicate that participants saw the risk as high, which would be expected to be greater with the foreground-only display), or (2) as evidence of *low* decrease in risk (and thus indicate that participants saw the decrease as low, which would be expected to be smaller with the foreground-only display). If participants were responding primarily the first way, we would expect (a) high correlations between perceived risk with and without the vaccine; and (b) positive correlations of a similar size between both variables with the perceived decrease in risk. If participants were responding primarily the second way, we would expect (a) low correlations between perceived risk with and without the vaccine; (b) a positive correlation between perceived risk without the vaccine and perceived decrease in risk; and (c) a negative correlation between perceived risk with the vaccine and perceived decrease in risk (since perceived decrease = perceived risk without the vaccine minus perceived risk with the vaccine). The results indicated (a) moderate to large correlations between perceived risk with and without the vaccine (r = .79 and .40 for risk reductions of 20% and 80%, respectively); (b) small positive correlations between perceived risk with the vaccine and perceived decrease in risk (r = .14 and .04); and (c) larger positive correlations between perceived risk without the vaccine and perceived decrease in risk (r = .42 and .59). These findings suggest that both of the above interpretations are at least partially correct, suggesting that the results with this measure are ambiguous.