Article:

Stone, ER, Bruine de Bruin, W orcid.org/0000-0002-1601-789X, Rogers, AM et al. (2 more authors) (2017) Designing Graphs to Communicate Risks: Understanding How the Choice of Graphical Format Influences Decision Making. Risk Analysis, 37 (4). pp. 612-628. ISSN 0272-4332

https://doi.org/10.1111/risa.12660

Running Head: GRAPHICAL RISK COMMUNICATION FORMATS

Designing Graphs to Communicate Risks:

Understanding How the Choice of Graphical Format Influences Decision Making

Manuscript accepted for publication in Risk Analysis

Eric R. Stone*,[1]

Wändi Bruine de Bruin[2,3]

Abigail M. Rogers[1]

Emily M. Boker[1]

Jacqueline MacDonald Gibson[4]

[1] Wake Forest University, Department of Psychology

[2] Leeds University Business School, Centre for Decision Research

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

[4] University of North Carolina at Chapel Hill, Department of Environmental Sciences and Engineering

*Corresponding author phone: tel: 336-758-5729; e-mail estone@wfu.edu

Abstract

Previous research suggests that the choice of graphical format for communicating risk information affects both understanding of the risk magnitude and the likelihood of acting to decrease risk. However, the mechanisms through which these effects work are poorly understood. To explore these mechanisms using a real-world scenario, we examined the relative impact of two graphical displays for depicting the risk of exposure to unexploded ammunition during potential land redevelopment. One display depicted only the foreground information graphically (a bar graph of the number of people harmed), and a second depicted the foreground and background graphically (a stacked bar graph representing both the number harmed and at risk). We presented 296 participants with either the foreground-only or the foreground and background graphical display and measured a broad set of outcome variables, examining (1) the graphical display effect on each of the outcome measures and (2) the pathways by which any display effects work to influence decision making. We found that the foreground-only graphical display increased perceived likelihood and experienced fear, which produced greater worry, which in turn increased risk aversion. In addition, a positive evaluation of the communication materials increased support for policies related to land redevelopment, whether those policies were risk-taking or risk-mitigating. Finally, the foreground-only graphical display decreased understanding of the risk magnitude, showing that approaches to accomplish one risk communication goal (promoting risk-averse decisions) may do so at the expense of another goal (increasing understanding).

Key Words: Risk Magnitudes, Display Formats, Graphs, Risk Communication, Unexploded Ammunition

1. Introduction

Throughout their lives, people face many important decisions that involve risks to their health, safety, finances, and overall well-being. An accurate understanding of the risks is necessary to make informed decisions regarding whether the risks are worth taking. To assist people in making these decisions, experts who possess relevant knowledge will often need to communicate their knowledge to the people making the decisions.[1-3] For example, doctors communicate the risks of medical treatments to their patients, and risk analysts communicate the risks of a potential public policy to voters. This paper focuses on one particularly important issue in risk communication: how best to convey the likelihood, or risk magnitude, of negative events occurring.

A particular challenge in communicating risk magnitudes is that many health and safety risks have a low risk magnitude, yet have serious potential consequences if they do occur.[4-8] One promising strategy entails the use of graphical displays (for reviews, see [2,3,9-13]). However, despite a substantial amount of research suggesting that graphical displays can be effective for accomplishing various risk communication goals,[14-21] there are "few best practices" (Lipkus[2], p. 709).

One of the primary reasons for this lack of best practices is that much of the research has been largely a-theoretical, and thus it is unclear why certain graphical displays have the effects they do.[2,11,13] Because most studies have focused on only one of many different risk communication goals (e.g., to increase comprehension, to promote risk-reduction behaviors, etc.),[2,22,23] most studies have used only one or a small number of outcome measures, and there is little consistency in the choice of outcome measures used in different studies.[24] As a result, it has often been difficult to evaluate the reasons for the particular results (e.g., what the more proximal effects are that lead to changes in decisions), as well as to draw systematic conclusions across studies varying substantially in both research stimuli and outcome measures.

For example, consider research that has investigated the impact of using a display that graphically depicts both the number of people affected (the numerator of the risk ratio, or 'the foreground') and the total number of people at risk (the denominator of the risk ratio, or 'the background') versus a display that graphically depicts only the number of people affected. One line of research has shown that, when communicating about events with low-probability risk magnitudes, foreground-only graphical displays (e.g., icon displays or bar graphs depicting the number of people affected) increase risk-averse behavior, both in comparison to a purely numerical display[(19,25-27)] and in comparison to graphical displays that depict both the foreground and background information, such as stacked bar graphs or pie charts.[(26,28)] On the other hand, displays that depict both the foreground and background graphically produce greater understanding of the risk magnitude than do foreground-only graphical displays.[(15)] As promoting risk-averse behavior and increasing understanding are two frequent goals of risk communication efforts,[(2,22,23)] this provides a problem for the risk communicator who is simultaneously trying to help people understand the risks and motivate them to be safer.

Unfortunately, however, no work has examined the relationship of these findings. Do foreground-only graphical displays increase risk aversion because of reduced understanding, or is this effect due to some other mechanism? Would this pattern of results even hold if the same stimuli had been used in both lines of research? Many details of the studies investigating risk aversion and understanding of the risk magnitudes were different, and in particular, the risk magnitudes used varied between the studies that measured risk aversion and understanding.[15,26,28] Indeed, previous work suggests that low-probability risks should be particularly influenced by the manner in which the information is displayed.[8,29, also see 30,31]

The goal of the present work is to compare the impact of using a foreground-only graphical display versus a foreground and background graphical display on a wide variety of decision-relevant outcome measures. So doing allows us not only to examine each effect in isolation but also to determine how any effects on decisions come about.

1.1 Redevelopment of land potentially contaminated with UXO

To explore these mechanisms using a real-world scenario, we focus on decisions concerning the redevelopment of one of the approximately 2,000 former U.S. military bases contaminated with unexploded ordnance (UXO), or ammunition that failed to explode during live weapons training.[32] Many of these sites have been transferred to civilian agencies for reuse, and the others are scheduled to do so in the future. However, because UXO may be unearthed and explode, these sites pose risks to construction workers involved in land redevelopment and to anyone else who enters the site.[33] Whether these risks are worth taking is an important public policy question facing the local community. To assist local residents with these types of decisions, MacDonald and colleagues[34,35] developed a simulation model to assess the spatial distribution of UXO after initial clean-up efforts. The applied aim of the present work was to determine how best to communicate this risk-magnitude information to support informed decision making, in order to develop the best possible communications directed towards communities affected by UXO sites.[36]

1.2 Outcome variables

In addition to understanding of the risk-magnitude information and risk aversion, we measured four other outcome variables. The first of these is one's belief about how large the risk magnitude is, which we refer to as perceived likelihood. Note that although understanding of the risk magnitude information and perceived likelihood both refer to one's cognitive interpretation of the risk-magnitude information, they are conceptually distinct. Understanding refers to the match between the actual risk magnitude and what the person thinks the risk magnitude is. In the present context, does the actual risk of exposure to UXO match what is in the person's head? Perceived likelihood, on the other hand, reflects how large the person sees the risk magnitude as being. One can misunderstand the risk magnitude by thinking the chance of exposure to UXO is greater than it really is (poor understanding and high perceived likelihood) or by thinking that the chance of exposure to UXO is less than it really is (poor understanding but low perceived likelihood).

People not only evaluate risks cognitively, but also have emotional reactions to the communications.(24,37) These emotions can be influenced by graphical displays of the information(38-39) (see Visschers et al.(40) for a review). Here, we examined whether displaying both the foreground and background graphically or just the foreground graphically produces greater experienced fear(41) while reading the information about unexploded ammunition.

Another important outcome variable is the *user's* evaluation of the communication materials.(42-44) If a display is disliked, for example, users may not attend to the information or just dismiss it.(9,43) As a result, they may be unwilling to make decisions regarding the risky event, since they may not feel confident regarding their understanding of the information being communicated.[45]

Finally, we measured the user's worry about exposure to UXO. Although worry is often considered to be an emotional reaction,[46] some research suggests that it is determined in part by cognitive factors.[47,48] In the present study, worry is distinct from experienced fear, as the latter concept refers to the participant's emotional reaction while reading the communication materials, whereas worry is an overall assessment of concern, in this context about potential land redevelopment.

1.3 Potential pathways from display effects to decision making

As discussed previously, there is a need to increase our understanding of not only what effects graphical features have, but also how these effects come about.[2,11,13] Therefore, we constructed a theoretical model of how graphical display formats influence decision making based on our analysis of the existing literature (see Figure 1) and then tested this model in the current work. According to our model, display formats directly affect perceived likelihood, experienced fear while reading the risk communication, understanding of the risk magnitudes, and user evaluation of the communication materials, which in turn influence worry about UXO, which in turn influences the eventual decisions. Note our model suggests that worry is not influenced directly by the graph format, but instead by other factors more proximal to graph format in the model. This follows from the work suggesting worry has both affective[46] and cognitive[47,48] determinants. For example, Baron et al.'s [47] results suggest that one's support of action to eliminate a risk is determined primarily by worry, and that worry in turn is determined mainly by beliefs regarding probabilities (but see Schade et al.[49] and Sjöberg[50] for contrasting results).

We expect both perceived likelihood and experienced fear will be positively associated with worry and ultimately the level of risk aversion exhibited in people's decisions.[38,47,51] However, the roles of understanding and user evaluation are less clear. The direction of the link with understanding is of particular interest and is likely context-dependent. For example, much medical research has assumed that risk-averse behavior is desirable (what Garcia-Retamero & Hoffrage,[17] p. 28, refer to as "beneficial risk avoidance") and thus implicitly that increasing understanding of the risks will lead people to (appropriately) take more risk-averse actions.[52] In other contexts, however, there is no generally accepted "correct" behavior. The current context is an example of this situation, as preferences for managing UXO risks depend on the individuals' risk tolerances and their evaluations of whether the benefits are worth the risks. Thus it is unclear if increasing understanding of the risks involved will increase, decrease, or have no effect on the overall level of risk aversion.[1]

Our model has at least two important implications for understanding the effects of graph format on our outcome measures. First, the direct influences of graph format on the proximal variables should be larger than their indirect influences on the more distal variables. Indeed, the distal variables of worry and decision making will be additionally influenced by factors not depicted in the model, such as general proclivity towards taking risks. Second, multiple pathways will produce effects on worry and decision making, and these pathways could either operate in tandem to produce stronger effects, or in opposition to produce no overall change in decision making. For example, presenting only the foreground graphically should increase the perceived size of the risk,[26,38] which would increase worry and risk aversion. However, if this display also decreases understanding, and understanding is associated with increased worry and risk aversion, then that pathway would produce reduced worry and risk aversion. In this case, the two pathways could counteract each other, producing no overall impact on decisions made. In sum, a full understanding of each of the individual context-specific pathways is necessary to determine what if any impact a display feature will have on decision making.

1.4 The present experiment

We presented participants with a graphical display depicting the risks of UXO exposure at a construction site on the former Fort Ord, California, that is slated for redevelopment and reuse by the local community. The display depicts risks to construction workers for each of four soil excavation digging depths. We varied whether the foreground or both the foreground and background were depicted graphically. In addition, we varied whether the risk magnitudes reflected the actual probability levels (between .01 and .10) or smaller probability levels (between .001 and .01), due to a concern that the actual probability levels would not be small enough to elicit display effects (see Shepperd et al.[30]). Our interest in probability level, then, focused predominantly on whether it produced an interaction with graph format. Finding no interaction would demonstrate that the display effects are robust as long as the probability levels are relatively low, whereas finding an interaction would show that the display effects are sensitive even to relatively small changes in probability level.

After seeing the graphical risk communication, participants completed a number of decision-making questions regarding construction at Fort Ord, as well as measures of their perceived likelihood, fear experienced while reading the risk communication, understanding of the risk-magnitude information, evaluation of the communication materials, and worry about the risky event. We examined (a) the effects of depicting only the foreground graphically versus both the foreground and background graphically on each of the outcome measures, and (b) the degree to which the model in Figure 1 captures the causal influences on decisions about UXO-related risks.

2. METHOD

2.1 Participants

Participants were 296 university students (160 female, 136 male). Participation provided one option to fulfill an introductory psychology course requirement. Most of the participants were unfamiliar with the subject matter. In particular, on average participants rated their prior knowledge of unexploded ammunition as 1.75 (on a 1-7 scale), and only 12 (4.1%) of the participants stated they knew what the letters “UXO” stand for. Nearly half of the participants (41.9%) of the participants stated that they had heard stories regarding incidents with unexploded ammunition, however.

2.2 Communication materials

Figure 2 shows the information that participants received at the beginning of the study. It began with a description of Fort Ord, an old military base that closed in 1994. Like most military bases, Fort Ord had a firing range that soldiers used for live weapons training. Some of the ammunition failed to explode, remaining in the ground today, with the potential to detonate at a later time. Participants were asked to imagine that they lived near Fort Ord and to consider new construction plans for this land.

To inform these decisions, participants received information regarding the risk of construction workers’ exposure to unexploded ammunition while digging at four different excavation depths. In all conditions both the foreground (number of people exposed to unexploded ammunition) and background (number of people at risk) were presented numerically.

In addition, we varied whether only the foreground (see Figure 2A) or both the foreground and background (see Figure 2B) were presented graphically, and whether the risk magnitude reflected the actual probabilities or smaller probabilities that were more in keeping with typical investigations of low-probability risks. The actual probabilities condition, shown in Figure 2A, entails the actual exposure likelihoods that were computed by MacDonald et al.[35] for each of four digging depths (0 out of 100, 6 out of 100, 7 out of 100, and 10 out of 100). Risk increases with digging depth, because increased depth corresponds to the need to excavate a larger volume of soil and hence a higher probability of encountering UXO. In the smaller probabilities condition, shown in Figure 2B, "100" was replaced by "1000," so that the exposure likelihoods were: 0 out of 1,000, 6 out of 1,000, 9 out of 1,000, and 10 out of 1,000. Graphically, this was reflected in the bars representing the foreground information being 1/10 the size of those in the actual-probabilities condition.

2.3 Outcome measures

We included six outcome variables to fully explore the effects of the presented graph formats: perceived likelihood, experienced fear, understanding, user evaluation of the communication materials, worry, and decision making. Each of these constructs has been operationalized in multiple ways in different studies. Our approach was to use multiple "component measures" of each construct, scoring these component measures in a manner as consistent as possible with the way they were previously used. We then transformed them if necessary for good psychometric properties, and averaged the component measures into an overall measure of the construct.[2] Although averaging across multiple measures loses some of the nuances captured by the individual measures of the construct (e.g., between assessing understanding in different ways), this approach provides a manageable set of outcome variables with improved reliability. We provide an overview of these measures below and in Table I. More details are available in the Supplementary Materials.

2.3.1 Perceived likelihood

We used six measures based on those used in previous studies.[(25,26,30,38)] Details on how each of these component measures were constructed are provided in the supplementary materials (pp. 3-5). Five measures used one-item Likert-type scales similar to those used in Shepperd et al.[(30)] and Stone et al.,[(26)] e.g., "What do you think is the overall chance of being exposed to unexploded ammunition when working at this building site?" In keeping with the approach taken by Chua et al.[(38)] and Stone et al.,[(25)] for the sixth measure we asked participants how many out of 450 construction workers would be exposed to unexploded ammunition at each of the four digging depths and then averaged and log-transformed those estimates (with higher numbers indicating greater perceived likelihood). Responses to the six component measures were then averaged (after all were z-scored) to produce an overall measure of perceived likelihood (Cronbach's $\alpha = .88$; all factor loadings > .52 in a one-factor solution).

2.3.2 Fear experienced while reading the risk communication

Participants indicated the extent to which they felt 12 different emotions while reading the information about unexploded ammunition. These items were adapted from Lerner et al.,[(41)] and each emotion was rated on a scale from 0 to 8, with higher numbers indicating stronger emotions. Here, we only report on the four items that comprise fear, which was assessed by asking participants to rate the extent to which they felt worried, fearful, frightened and terrified. We averaged these four items into an overall measure of experienced fear (Cronbach's α = .95; all factor loadings > .89 in a one-factor solution).

2.3.3 Understanding of the risk magnitude information

We assessed understanding using seven component measures adapted from previous research.[25,38,45,53-56] Participants answered these items without being allowed to refer to the communication materials, thus preventing mindless copying of the presented information. Details on how each of these component measures were constructed are provided in the supplementary materials (pp. 12-17).

To summarize, we included two measures of understanding of absolute risk magnitudes,[25,38,45] which assess participants' ability to recall the absolute magnitudes of the likelihood information correctly. In keeping with the recommendations of previous research,[25,57-60] we attempted to capture how accurately people could recall both the verbatim (literal) information as well as its gist (meaning). Specifically, in keeping with the approach of Stone et al.,[25] we both asked verbatim recall questions at each digging depth (e.g., "I think that ___ out of ____ workers at this building site will be exposed to unexploded ammunition when digging one foot into the ground."), as well as questions using a denominator different from that used in the communication (asking how many out of 450 construction workers would be exposed to unexploded ammunition) to better capture participants' gist by making it more difficult to respond with the verbatim information. For each of these sets of items, we scored correctness as how close the estimates were to the actual risk magnitudes, and then averaged these correctness values.

Note that, as discussed above, the "out of 450" items were also used to assess perceived likelihood, but were scored here as closeness to the correct values rather than as the size of the estimates.

We also included four measures of understanding of relative risk magnitudes, which assess participants' ability to understand relative relationships among the likelihood of exposure at the four digging depths. In keeping with the work of Stone et al.,[(25)] two of these measures were based on the items used in measuring understanding of absolute risk magnitudes discussed above. Specifically, since we had estimates at each of the digging depths, we correlated these estimates with the true values, thus assessing how accurate participants perceived the relative orderings of the risk magnitudes. The third measure was based on the work of Cuite et al.[(53)] and evaluated participants' ability to perform tradeoff operations on the risk-magnitude information by requiring them to tradeoff risk and time at different digging depths. For example, participants were asked whether it was riskier to dig 6 inches for one month and 4 feet for a second month or at 1 foot for both months. The fourth measure was based on the work of Weinstein[(55)] and evaluated participants' understanding regarding which change in digging depth (e.g.., going from no digging to 6 inches of digging or from 6 inches to 1 foot of digging) would produce the greatest increase in risk.[3]

Finally, we included one measure of subjective understanding.[(56)] Specifically, we asked participants "How well did you understand the information you were just given about how many people would be exposed to unexploded ammunition during construction on this building site?"

Despite the fact that we assessed understanding via seven different component measures, a one-factor solution explained a high percentage (47.2%) of their variance. Thus, we z-scored and averaged the seven different component measures to create an overall measure of understanding (Cronbach's α = .80; all factor loadings > .34 in a one-factor solution).

2.3.4 User evaluation of the communication materials

We used five items adapted from previous research[42,44,45] to assess how participants evaluated the risk communication. A sample question is: "How much did you like the way the information regarding the chance of exposure was presented to you?" (1 = did not like it, and 7 = liked it a lot). Participants rated the information on the dimensions *helpful, accurate, useful, credible* and *liking* on a scale from 1 to 7, with higher scores indicating a more positive evaluation. These items were averaged into an overall user evaluation score (Cronbach's α = .85; all factor loadings > .69 in a one-factor solution).

2.3.5 Worry about exposure to UXO

Nine items were adapted from previously used scales.[46,56] These items assessed how worried, afraid and anxious participants would be to work at the construction site, live near it or let children play near the building site on a 7-point Likert-type scale. A sample question asked: "If you were a construction worker at this building site, how worried would you be about finding unexploded ammunition during your work there?" Responses to these items were averaged into an overall measure of worry (Cronbach's α = .93; all factor loadings > .72 in a one-factor solution).

2.3.6 Decisions regarding land redevelopment

Twelve items adapted from Bruine de Bruin et al.[45] asked participants to make decisions about aspects of the construction. In keeping with the results of Bruine de Bruin et al.,[45] a factor analysis revealed that these items loaded onto two factors, which we labeled support for construction activities and support for increased pay and education for construction workers. The support for construction activities factor consists of six items that assessed the willingness of participants to support, work at or live near the construction site on a 7-point Likert-type scale, with higher scores indicating greater support for the construction. A sample question is: "If this were your decision, would you authorize construction to proceed at this building site?" These six items were averaged into an overall level of support for construction activities (Cronbach's $\alpha = .91$; all factor loadings > .68 after rotation of a two-factor solution).

The support for increased pay and education for construction workers factor consists of six items that asked participants to indicate the extent to which they would support the use of educational materials for workers at the construction site and request hazardous pay for working there. A sample question asked: "As a construction worker at this building site, would you attend an optional educational program concerning unexploded ammunition safety?" All items were on a 7-point Likert-type scale, with higher numbers indicating greater support for the risk-mitigating factors of educational materials and hazardous pay. These six items were averaged into an overall level of support for increased pay and education for construction workers (Cronbach's $\alpha = .78$; all factor loadings > .59 after rotation of a two-factor solution).

2.4 Covariates

We asked four questions regarding variables that might covary with our dependent measures: gender, knowledge about unexploded ammunition prior to the study, understanding what the letters "UXO" stand for, and having heard any stories about incidents with unexploded ammunition.

2.5 Procedure

Participants were randomly assigned to one of four conditions, varying whether or not the background was presented graphically and the size of the risk magnitudes. All participants first spent three minutes reading over the communication material as described above. They then placed the sheet face down and were not allowed to return to it for the remainder of the study, so as to prevent them from simply copying answers to understanding questions directly from the displays. Next, they were asked to spend two minutes reading information about the town of Monterey, California, which is near Fort Ord. This information provided participants with a context for their subsequent decisions and made it more difficult for them to recall the precise information provided.

After reading these materials, participants completed a questionnaire that contained the outcome measures. Items for some of the constructs (such as understanding) were distributed throughout the questionnaire. In addition, we asked questions with a greater potential for diminishment over time (such as the questions about negative emotions experienced while reading the risk communication) earlier, and items with a potential to influence other questions (such as questions that entailed precise numbers) later in the survey. Finally, participants completed the covariate measures.

2.6 Analytic plan

For each of the overall outcome measures we conducted a-priori 2 (Graph Format: foreground-only vs. foreground and background) by 2 (Probability Level: out of 100 [actual probabilities] vs. out of 1000 [smaller probabilities]) ANCOVAs, covarying out gender, UXO knowledge, UXO familiarity, and exposure to UXO stories. The results for the main effect of graph format are summarized in Table II. Analyses on the component measures are provided in the supplementary material, with findings of particular interest mentioned briefly here. We then examined how well the model in Figure 1 explains the pattern of our results using structural equation modeling, adjusting the model in an exploratory manner to account for relationships among the variables that differed from our original model.

3. RESULTS

3.1 Analyses examining the effects of graph format

3.1.1 Perceived likelihood

There were main effects of both Graph Format and Probability Level on the size of the perceived risk magnitude. Specifically, participants presented with the actual probability levels (M z-score = .19) judged the risk to be more likely than did participants presented with the smaller probability levels (M z-score = -.19), $F(1, 288) = 20.37$, $p < .001$. More importantly, participants presented with the foreground-only graphical display (M z-score = .12) judged the risk to be more likely than did participants presented with the foreground and background graphical display (M z-score = -.12), $F(1, 288) = 11.70$, $p = .001$. There was no interaction between Graph Format and Probability Level, $F(1, 288) = .52$, $p = .47$.

3.1.2 Fear experienced while reading the risk communication

There was a main effect of Graph Format, whereby participants presented with the foreground-only graph (M = 5.08) reported experiencing more fear when reading the communication materials than did participants presented with the foreground and background graph (M = 4.32), $F(1, 288) = 10.03$, $p = .002$. Neither the main effect of

Probability Level nor the interaction was significant (both F's < 1).

3.1.3 Understanding of the risk magnitude information

There was a main effect of Graph Format and a marginal effect of Probability Level on participants' overall understanding of the risk-magnitude information. Participants presented with the foreground and background graph (M z-score = .09) understood the risk information better than did participants presented with the foreground-only graph (M z-score = -.07), $F(1, 279) = 6.89$, $p = .009$.[4] In addition, participants presented with the actual probability levels (M z-score = .08) understood the risk information marginally better than did participants presented with the smaller probability levels (M z-score = -.05), $F(1, 279) = 3.35$, $p = .07$. There was no interaction between Graph Format and Probability Level, $F(1, 279) = .55$, $p = .46$.

3.1.4 User evaluation of the communication materials

There was no main effect of Graph Format, Probability Level, or interaction between them on how much participants liked the information presented to them (all p's > .22).

3.1.5 Worry about exposure to UXO

There was a main effect of Probability Level, whereby participants worried more in the actual probabilities condition (M = 5.65) than in the smaller probabilities condition (M = 5.34), $F(1, 288) = 6.04$, $p = .01$. Although there was a trend for participants presented with the foreground-only graph (M = 5.58) to worry more than participants presented with the foreground and background graph (M = 5.41), this difference did not reach significance, $F(1, 288) = 1.81$, $p = .18$.[5] The interaction between the variables did not approach significance, $F(1, 288) = .01$, $p = .93$.

3.1.6 Decisions regarding land redevelopment

There was a marginally significant main effect of Graph Format on support for construction activities, whereby participants presented with the foreground-only graph (M = 2.35) were less in favor of the construction activity than were participants presented with the foreground and background graph (M = 2.59), $F(1, 288) = 3.31$, $p = .07$. There were no other effects on either support for construction activities or on support for increased pay and education for construction workers, all p's > .39.

3.2 Contributors to decisions regarding land redevelopment

As discussed previously, we are interested not only in determining the effects on each variable individually, but also in examining how any effects on decision making come about. Table III provides the zero-order correlations among the primary outcome measures. As is evident from the large correlations in the table, the primary determinant of both types of decisions was the amount of worry. This supports our conjecture earlier that, for situations involving risk, amount of worry is the most proximal indicator of decision making, as shown in Figure 1.

3.2.1 Modeling the effect of graph format on support for construction activities

We tested the ability of the model in Figure 1 to predict support for construction activities decisions, eliminating user evaluation of the communication materials, since that variable was not influenced by graph format. In particular, we constructed a structural equation model including all of the links depicted in Figure 1 (except for user evaluation), as well as covariances among understanding, perceived likelihood, and experienced fear,[6] to determine whether the links indicated in that model were necessary. All of the links were significant at $p < .05$, except for the link between understanding and worry, standardized beta = .06, $p = .23$. We thus eliminated that link from the model.

Next, we tested whether any of the direct links between the variables on the left-hand side of Figure 1 and support for construction activities would add predictive validity by adding links one at a time between graph format → support for construction activities, perceived likelihood → support for construction activities, experienced fear → support for construction activities, and understanding → support for construction activities, in keeping with the logic used in stepwise procedures in regression analysis. None of these links was significant, all p's > .26. Our final model of the effect of graph format on support for construction activities is depicted in Figure 3, $\chi^2(6) = 4.42$, p =.62; RMSEA = .000; NFI = .989. The non-significant chi-square, low RMSEA (Root Mean Square Error of Approximation), and high NFI (Normed Fit Index) all indicate good fits of our model.[61] In summary, graph format influenced perceived likelihood and experienced fear while reading the communication materials, which influenced worry, which in turn influenced support for construction activities. Graph format also influenced understanding, but participants' understanding of the risk-magnitude information was unrelated to worry or their support for construction activities. Further, there were no direct links between either graph format, perceived likelihood, experienced fear, or understanding with support for construction activities.

3.2.2 Modeling the effect of graph format on increased pay and education

We then took the same approach, this time predicting support for increased pay and education for construction workers. Although there was no effect of graph format on this variable, examining the structural equations model provides some indication of why this was the case. Our final model for predicting support for increased pay and education for construction workers is also depicted in Figure 3, and the fit indices again indicate a good fit of our model, $\chi^2(5) = 4.93$, p =.43; RMSEA = .000; NFI = .982. There were two primary differences between this model and the previous one. First, the worry → decision making link (standardized beta = .42) was considerably weaker than in the model predicting support for construction activities (standardized beta = -.67). Second, there was a significant direct relationship between understanding of the risk-magnitude information and support for increased pay and education, standardized beta = .17, p = .002.

These two differences explain why there was no effect of graph format on support for increased pay and education for construction workers. First, the relationship with worry was weaker for increased pay and education. Second, the two pathways from graph format to increased pay and education went in opposite directions. On the one hand, the foreground-only graph increased the perceived likelihood of harm and experienced fear, which increased worry and thus support for increased pay and education. On the other, the foreground-only graph decreased understanding, and decreasing understanding led to reduced support for increased pay and education. Thus, the addition of the understanding – support for increased pay and education link in Figure 3 served to reduce the impact of graph format that occurred through the perceived likelihood and experienced fear variables.

3.2.3 Other influences on land redevelopment decisions

Although the primary goal of our path analyses was to determine how the choice of graph format influences downstream decision making, understanding the factors (whether produced by the graph or by other variables) that influence decisions is important in its own right. An examination of the results shown in Figure 3 and Table III shows an intriguing difference behind how different factors influence decision making. First, perceived likelihood, experienced fear, and worry were negatively related to support for construction activities and positively related to support for increased pay and education. These relationships make sense, in that construction is a risk-taking activity and increasing pay and education is a risk-mitigating activity.

However, as seen in Table III, a positive evaluation of the communication materials was positively correlated with both support for construction activities and support for increased pay and education for construction workers, although not significantly with support for construction activities. To determine if these relationships held with the rest of the variables in the model, we took the full model depicted in Figure 3 and added links between user evaluation and each of the decision-making variables (also including the covariances between user evaluation and perceived likelihood, experienced fear, and understanding). The relationships between user evaluation and both decision-making variables were positive and either significant or marginally significant, standardized beta = .16, $p = .002$ for support of increased pay and education for constructions workers, and standardized beta = .07, $p = .10$ for support of the construction activities. Thus, having a positive evaluation of the communication materials increased people's support for different types of policies, whether risk-taking or risk-mitigating.

4. DISCUSSION

We evaluated how best to communicate risk-magnitude information in the context of decisions about the redevelopment of land potentially contaminated with unexploded ordnance (UXO) by comparing a display that graphically depicted only the foreground to a display that graphically depicted both the foreground and background. Below, we discuss our findings pertaining to graph format effects on decision making and related outcomes, as well as pertaining to the mechanisms by which the choice of graphical format affects decision making.

4.1 Graph format effects

In comparison to those presented with the foreground and background graph format, participants presented with the foreground-only graph format 1) perceived the likelihood of exposure to UXO to be greater, 2) experienced more fear when reading the risk communication materials, 3) understood the risk-magnitude information less well, and showed trends to 4) worry more and 5) be less apt to support construction activities.

Of particular importance is that the graphical display that led to the greatest perceived risk magnitude, experienced fear and anger, and more risk-averse decisions also produced the least understanding of the risk magnitudes. Evidently, foreground-only graphical displays are effective at accomplishing one frequent goal of risk communication efforts (increasing perceived likelihood and risk aversion) but do so at the expense of another important goal (increasing understanding of the risk information) (see also [(25)]).

4.2 Identifying the pathways by which graphical displays influence decisions

As discussed by Lipkus and Hollands,[(11)] it is important to understand not just what effects graphical displays have, but also how these effects occur. To this end, we examined the determinants of participants' decisions and the role of graphical displays in producing changes in them. There were four main findings from this set of analyses.

First, the primary determinant of both types of decisions, support for construction activities and support for increased pay and education for construction workers, was the amount of worry. As seen in Table III and Figure 3, the link between worry and subsequent decisions was much stronger than that between any of our other outcome variables and the decisions made. This finding is consistent with the work of Baron et al.,[(47)] who found that worry was the main determinant of action priority.

Second, one way by which graphical displays influence decision making is by influencing how large people perceive the risk to be and their experienced fear while reading the communication materials, which in turn influence worry, and subsequently decision making. Similarly, Baron et al.[(47)] found that concern for action is strongly related to worry and that worry, in turn, is determined largely by probability beliefs. Our results extend Baron et al.'s[(47)] conclusion by demonstrating that both emotions and the perceived size of the risk magnitudes determine worry, and that worry mediates the effects of both these variables on subsequent decision making.

Third, in contrast to the results with perceived likelihood and experienced fear, we found that liking the communication material influenced decision making through a causal pathway that did not include worry. Interestingly, whereas perceiving the risk magnitude to be large and experiencing negative emotions *increases one's* risk-averse behavior generally, having a positive evaluation of the communication materials i*ncreases one's* tendency to support decision-making policies, whether these policies entail risk taking (supporting the construction activities) or risk mitigation (supporting increased pay and education for construction workers). Note that, due to the correlational nature of the current findings, it is possible that the directional relationship is reversed, in that people who generally support decision-making policies are more apt to like the risk communication.[7] Previous work we have conducted,[(45)] however, showed that the effect of a communication display on decision making was mediated by liking of the materials. Combining that finding with the current results, we posit that people who like the communication materials feel more capable of taking action on the basis of that information, both because of greater self-efficacy about decisions and because of greater trust in the competence of the people carrying out the actions. Although these effects are relatively small in relation to our other documented effects on decision making, they suggest that one relatively straightforward way of encouraging people to take action is simply to design communication materials that they like.

Fourth, although the foreground-only graphical display decreased understanding while increasing perceived likelihood and experienced fear, only the latter two variables (and not understanding) were significantly related to support for construction activities. Thus, it appears that people presented with this display were unwilling to support construction at the site due to their increased worry, rather than to a decreased understanding of the risk magnitude.

4.3 Boundary conditions

Further research is needed to determine whether or not the results generalize to other situations and participants, but there are four issues that warrant discussion in particular: 1) the specific context involved, 2) the size of the probabilities, 3) the choice of information to depict graphically, and 4) the participants in our study.

As discussed above, people's decisions were determined primarily by the amount of worry, not by their understanding of the risk magnitudes. This lack of relationship with understanding in this study makes sense, since there is no universally "correct" decision about whether to allow construction to move forward given the exposure risks posed in this study. Rather, preferences for managing UXO risks depend on the individual's risk tolerance – perhaps even more so for UXO risks than for many other types of risks because there is a great deal of uncertainty regarding the consequences of digging up UXO.[62] In other situations, however, an accurate understanding of the risk magnitudes may be associated with specific decisions, such as being risk-averse in certain health scenarios.[52]

We found no interactions with probability level, suggesting that our effects of graph format held equivalently for the actual risk magnitudes and for those of an order of magnitude lower. However, to be realistic in the present context both sets of probabilities were at the “moderately small” level. As previous work has found that the effectiveness of specific graphical displays varies according to the size of the risk magnitudes,[29,63] an important question for future research is whether this pattern of results would hold for substantially larger probabilities.

Further, our finding that foreground-only graphical displays increased risk aversion is a function of the fact that these displays were communicating the risk of digging. If instead the benefits of digging were depicted with foreground-only graphical displays, so doing would presumably increase the perceived likelihood of the benefits and thus decrease risk aversion. Also, in many cases multiple types of risks are involved, for example, a patient with clogged arteries might consider taking a drug that would reduce the likelihood of requiring bypass surgery but have the risk of other side effects.[57] In these situations, it seems likely that both risks would be overestimated with foreground-only graphical displays.

Finally, the participants in our study were college students, not residents of a community surrounding the Fort Ord UXO site. Although the current work did not study that participant population, it is consistent with other research we have conducted with that community.[36] In that research, we found that, in comparison to only textual information, a display that depicted the foreground and background graphically increased understanding of the risk magnitudes by the Fort Ord residents. Although that work did not include a foreground-only graphical display, it is consistent with the current findings in that both studies suggest that foreground and background graphical displays help participants to gain an accurate understanding of the risk magnitudes. If anything, we expect that any effects of using college students would have been to dampen our results, given college students' relatively high levels of numeracy, and because other research[25] shows that foreground-only graphical displays lead lower numeracy participants in particular to overestimate the risks.

4.4 Implications

Two important risk-communication goals are to increase understanding of the risks and to increase risk aversion.[2,22,23] Unfortunately, our work clearly shows that a graphical display that is effective at meeting one of these goals may be ineffective for meeting the other goal. Participants presented with a graphical display depicting both the foreground and background graphically may dismiss the risks out of hand, even when public health experts would deem it beneficial for them to implement risk-reduction behaviors (see Sheppard et al.[30] for a discussion of this issue). Thus, foreground-only graphical displays may be a better choice for improving public safety. However, our work also suggests that interventions solely designed to increase risk aversion may come at a cost, decreasing people's understanding of the actual risk magnitudes involved – in essence disempowering them from making well-informed decisions.

More generally, our work suggests there are no easy answers for deciding how the ethical risk communicator should convey the risk information. However, by determining what factors influence people's decisions and evaluating the effects of the potential intervention on each of these factors, it is possible to understand not just what effects a communication technique is having but also why these effects occur. Armed with this knowledge, the risk communicator is then in the optimal position to weigh the benefits and costs of the communication approach.[64]

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (#SES 0922315).

We thank Hannah Stroup for her help in formatting the manuscript.

REFERENCES

1. Fischhoff, B. Risk perception and communication unplugged: Twenty years of process. Risk Analysis. 1995;15:137-145.
2. Lipkus, I. M. Numerical, verbal and visuals formats of conveying health risks: Suggested best practices and future recommendations. Medical Decision Making. 2007;27:696-713.
3. Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & de Vries, N. N. K. Probability information in risk communication: A review of the research literature. Risk Analysis. 2009;29:267-287.
4. Camerer, C. R, & Kunreuther, H. Decision processes for low probability events: Policy implications. Journal of Policy Analysis and Management. 1989;8:565-592.
5. Fisher, A., McClelland, G. H., & Schulze, W. D. Communicating risk under Title III of SARA: Strategies for explaining very small risks in a community context. Journal of Air Pollution Control Association. 1989;39:271-276.
6. Halpern, D. F., Blackman, S., & Salzman, B. Using statistical risk information to assess oral contraceptive safety. Applied Cognitive Psychology. 1989;3:251-260.
7. Kunreuther, H., Novemsky, N., & Kahneman, D. Making low probabilities useful. The Journal of Risk and Uncertainty. 2001;23:103-120.
8. Stone, E. R., Yates, J. F., & Parker, A. M. Risk communication: Absolute versus relative expressions of low-probability risks. Organizational Behavior and Human Decision Processes. 1994;60:387-408.
9. Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. Design features of

graphs in health risk communication: A systematic review. Journal of the American Medical Informatics Association. 2006;13:608-618.

10. Garcia-Retamero, R., & Cokely, E. T. Communicating health risks with visual aids. Current Directions in Psychological Science. 2013;22:392-399.
11. Lipkus, I. M., & Hollands, J. The visual communication of risk. Journal of the National Cancer Institute Monographs. 1999;25:149-163.
12. Spiegelhalter, D., Pearson, M., & Short, I. Visualizing uncertainty about the future. Science. 2011;333:1393-1400.
13. Trevena, L. J., Zikmund-Fisher, B. J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P. K. J., King, J., Lawson, M. L., Linder, S. K., Lipkus, I., Ozanne, E., Peters, E., Timmermans, D., & Woloshin, S. Presenting quantitative information about decision outcomes: A risk communication primer for patient decision aid developers. BMC Medical Informatics and Decision Making. 2013;13(Suppl 2):1-15.
14. Fagerlin, A., Wang, C., & Ubel, P. A. Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? Medical Decision Making. 2005;25:398-405.
15. 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.
16. Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. Do icon arrays help reduce denominator neglect? Medical Decision Making. 2010;30:672-684.
17. Garcia-Retamero, R., & Hoffrage, U. Visual representation of statistical

information improves diagnostic inferences in doctors and their patients. Social Science & Medicine. 2013;83:27-33.

18. Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. Journal of Behavioral Decision Making. 2012;25: 390-401.
19. Stone, E. R., Yates, J. F., & Parker, A. M. Effects of numerical and graphical displays on professed risk-taking behavior. Journal of Experimental Psychology: Applied. 1997;3:243-256.
20. Tait, A. R., Voepel-Lewis, T., Zikmund-Fisher, B. J., & Fagerlin, A. The effect of format on parents' understanding of the risks and benefits of clinical research: A comparison between text, tables, and graphics. Journal of Health Communication. 2010;15:487-501.
21. Zikmund-Fisher, B. J., Ubel, P. A., Smith, D. M., Derry, H. A., McClure, J. B., Stark, A., Pitsch, R. K., & Fagerlin, A. Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. Patient Education and Counseling. 2008;73:209-214.
22. Keeney, R. L., & von Winterfeldt, D. Improving risk communication. Risk Analysis. 1986;6:417-424.
23. Rohrmann, B. The evaluation of risk communication effectiveness. Acta Psychologica. 1992;81:169-192.
24. Edwards, A., & Elwyn, G. How should effectiveness of risk communication to aid patients' decisions be judged? A review of the literature. Medical Decision Making. 1999;19:428-434.

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

26. Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. Organizational Behavior and Human Decision Processes. 2013;90:19-36.

27. Schirillo, J. A., & Stone, E. R. The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. Risk Analysis. 2005;25:555-566.

28. 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.

29. Stone, E. R., & Rush, C. J. Risk communication: The effectiveness of graphical modes depends on the risk magnitude. Poster presented at: Society for Judgment and Decision Making Annual Meeting; 1997, November; Philadelphia, PA.

30. Shepperd, J. A., Lipkus, I. M., Sanderson, S. C., McBride, C. M., O'Neill, S. C., & 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.

31. LeClerc, J., & Joslyn, S. Odds ratio forecasts increase precautionary action for extreme weather events. Weather, Climate & Society. 2012;4:263-270.

32. Department of Defense. (Department of Defense, Washington, D.C.) Defense

environmental restoration program annual report to Congress, Fiscal Year 2005 [Internet]. Available from: https://www.denix.osd.mil/denix/Public/News/OSD/DEP2005/deparc2005.html

33. Environmental Protection Agency. Draft handbook on the management of ordnance and explosives at closed, transferred, and transferring ranges. Washington, D.C.: EPA, Office of Solid Waste and Emergency Response; 2001.
34. MacDonald, J. A., & Small, M. J. Assessing sites contaminated with unexploded ordnance: Statistical modeling of ordnance spatial distribution. Environmental Science and Technology. 2006;40:931-938.
35. MacDonald, J. A., Small, M. J., & Morgan, M. G. Quantifying the risks of unexploded ordnance at closed military bases. Environmental Science and Technology. 2009;43:259-265.
36. MacDonald Gibson, J., Rowe, A., Stone, E. R., & Bruine de Bruin, W. Communicating quantitative information about unexploded ordnance risks to the public. Environmental Science & Technology. 2013;47:4004-4013.
37. Peters, E., Lipkus, I., and Diefenbach, M. A. The functions of affect in health communications and in the construction of health preferences. Journal of Communication. 2006;56:S140-S162.
38. Chua, H. F., Yates, J. F., & Shah, P. Risk avoidance: Graphs versus numbers. Memory & Cognition. 2006;34:399-410.
39. Timmermans, D. R. M., Ockhuysen-Vermey, C. F., & Henneman, L. Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluations and decisions. Patient Education and Counseling.

2008;73:443-447.

40. Visschers, V. H. M., Wiedemann, P. M., Gutscher, H., Kurzenhäuser, S., Seidl, R., Jardine, C. G., & Timmermans, D. R. M. Affect-inducing risk communication: Current knowledge and future directions. Journal of Risk Research. 2012;15:257-271.

41. Lerner, J. S., Gonzalez, R. M., Small, D. A., and Fischhoff, B. Effects of fear and anger on perceived risks of terrorism: A national field experiment. Psychological Science. 2003;14:144-150.

42. Dolan, J. G., & Iadarola, S. Risk communication formats for low probability events: An exploratory study of patient preferences. BMC Medical Informatics and Decision Making. 2008;8:1-9.

43. Fortin, J. M., Hirota, L. K., Bond, B. E., O'Connor, A. M., & Col, N. F. Identifying patient preferences for communicating risk estimates: A descriptive pilot study. BMC Medical Informatics and Decision Making. 2001;1. Available from http://www.biomedcentral.com/1472-6947/1/2

44. Schapira, M. M., Nattinger, A. B., & McAuliffe, T. L. The influence of graphic format on breast cancer risk communication. Journal of Health Communication. 2006;11:569-582.

45. Bruine de Bruin, W., Stone, E. R., MacDonald, J., Fischbeck, P., & Shoraka, M. B. The effect of communication design and recipients' numeracy on responses to UXO risk. Journal of Risk Research. 2013;16:981-1004.

46. Lipkus, I. M., Klein, W. M. P., Skinner, C. S., & Rimer, B. K. Breast cancer risk perceptions and breast cancer worry: What predicts what? Journal of Risk

Research. 2005;8:439-452.

47. Baron, J. Hershey, J. C., & Kunreuther, H. Determinants of priority for risk reduction: The role of worry. Risk Analysis. 2000;20:413-427.
48. MacGregor, D. Worry over technological activities and life concerns. Risk Analysis. 1991;11:315-324.
49. Schade, C., Kunreuther, H., & Koellinger, P. Protecting against low-probability disasters: The role of worry. Journal of Behavioral Decision Making. 2012;25:534-543.
50. Sjöberg, L. Worry and risk perception. Risk Analysis. 1998;18:85-93.
51. DeFrank, J. T., Carey, L. A., & Brewer, N. T. Understanding how breast cancer patients use risk information from genomic tests. Journal of Behavioral Medicine. 2013;36:567-573.
52. Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. Aversion to side effects in preventive medical treatment decisions. British Journal of Health Psychology. 2007;12:383-401.
53. Cuite, C. L., Weinstein, N. D., Emmons, K., & Colditz, G. A test of numeric formats for communicating risk probabilities. Medical Decision Making. 2008;28:377-384.
54. Miron-Shatz, T., Hanoch, Y., Graef, D., & Sagi, M. Presentation format affects comprehension and risk assessment: The case of prenatal screening. Journal of Health Communication. 2009;14:439-450.
55. Weinstein, N. D. What does it mean to understand a risk? Evaluating risk comprehension. Journal of the National Cancer Institute Monographs. 1999;25:15-

20.

56. Wright, A. J., Whitwell, S. C. L., Takeichi, C., Hankins, M., & Marteau, T. M. The impact of numeracy on reactions to different graphic risk presentation formats: An experimental analogue study. British Journal of Health Psychology. 2009;14:107-125.
57. Hawley, S. T., 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.
58. Reyna, V. F. A theory of medical decision making and health: Fuzzy trace theory. Medical Decision Making. 2008;28:850-865.
59. Reyna, V. F., & Brainerd, C. J. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learning and Individual Differences. 2008;18:89-107.
60. Zikmund-Fisher, B. J. The right tool is what they need, not what we have: A taxonomy of appropriate levels of precision in patient risk communication. Medical Care Research and Review. 2013;70 (1 suppl):37S-49S.
61. Wuensch, K. L. Conducting a path analysis with SPSS/AMOS [Internet lecture notes]. 2008. Available from: http://core.ecu.edu/psyc/wuenschk/MV/SEM/Path-SPSS-AMOS.pdf .
62. MacDonald, J. A., Small, M. J., & Morgan, M. G. Explosion probability of unexploded ordnance: Expert beliefs. Risk Analysis. 2008;28:825-841.
63. McCaffery, K. J., Dixon, A., Hayen, A., Jansen, J., Smith, S., & Simpson, J. M. The influence of graphic display format on the interpretations of quantitative risk

information among adults with lower education and literacy: A randomized experimental study. Medical Decision Making. 2012;32:532-544.

64. Witte. K. The manipulative nature of health communication research: Ethical issues and guidelines. American Behavioral Scientist. 1994;38:285-293.
65. Carrigan, N., Gardner, P. H., Conner, M., & Maule, J. The impact of structuring information in a patient decision aid. Psychology and Health. 2004;19:457-477.
66. Holmes-Rovner, M., Kroll, J., Rovner, D. R., Schmitt, N., Rothert, M., Padonu, G., & Talarczyk, G. Patient decision support intervention: Increased consistency with decision analytic models. Medical Care. 1999;37:270-284.
67. O'Connor, A. M., Tugwell, P., Wells, G. A., Elmslie, T., Jolly, E., Hollingworth, G., McPherson, R., Bunn, H., Graham, I., & Drake, E. A decision aid for women considering hormone therapy after menopause: Decision support framework and evaluation. Patient Education and Counseling. 1998;33:267-279.
68. Simmons, J. P., Nelson, L. D., & Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science. 2011;22:1359-1366.

Footnotes

[1] It is worth emphasizing that even if there is no impact on the *average* decision made by increasing understanding, there could still be improvements in decision making. For example, people might not be generally more risk averse or risk seeking, but make decisions more in line with their values,[(65-67)] which could in some cases produce more risk- averse and in other cases more risk-seeking decisions. Thus, increasing understanding is an important risk communication goal, even if it does not produce a mean shift in the decisions made.

[2] Importantly, all scoring decisions were made in advance of conducting any inferential analyses and without knowledge of condition, thus ensuring that we did not capitalize on chance in our inferential analyses.[(68)] For similar reasons, we used our overall measure of each construct in our main analyses, rather than the subset of items that was most significant.

[3] We also asked participants to directly rank the risks, but since almost all of our participants (94%) got these rankings completely correct, we did not include that variable due to the ceiling effect. Note that all the measures of understanding of relative risk magnitude that we included required participants to understand the extent of differences between the digging depths (not just the ordering), which produced more variability in these variables.

[4] Although there was an overall effect of graph format on understanding, this effect was particularly strong for measures of understanding of absolute risk magnitudes, in keeping with the results of Stone et al.[(25)] See the supplementary materials (pp. 17-19) for more details.

[5] Although the display effect on overall worry was non-significant, when we included only the three items regarding working at the building site, those participants presented with the foreground-only graphical display (M = 5.91) were more worried than those given the display that depicted both the foreground and background graphically (M = 5.54), F (1, 288) = 6.96, p = .009. That the effect was stronger on these items makes sense, in that the risk-magnitude information provided to the participants was in regards to working at the building site, rather than about children playing at the site or living near Fort Ord. See the supplementary materials (pp.25-27) for more details.

[6] We decided a priori to maintain all the covariances among perceived likelihood, experienced fear, and understanding regardless of their significance levels, since these links were not of theoretical interest and were included only to provide a more accurate final model. In keeping with the zero-order correlations presented in Table III, both links with perceived likelihood were highly significant (p's < .001), but the covariance between experienced fear and understanding of the risk magnitudes was not significant, covariance = -.07, p = .40. Note that the perceived likelihood – experienced fear relationship is intuitive and seems likely to hold in most studies. In contrast, the link between perceived likelihood and understanding is likely unique to the current context, as most participants who did not accurately estimate the risks overestimated them in our study, producing the negative relationship.

[7] We than an anonymous reviewer for mentioning this possibility.

Author Notes

Emily Boker is now a Licensed Professional Counselor at Lawrence Hall.

Table I

Description, coding, and descriptive statistics for the study's primary outcome variables

Measure	Description	Coding	Confidence Interval	Skewness	Kurtosis
Perceived likelihood	Mean of 6 component measures, standardized	Standardized Higher=Greater Perceived Likelihood	[-.09, .09]	.18	-.45
Experienced fear	Mean of 4 items: worried, fearful, frightened, terrified	Scale from 0-8 Higher=Greater Fear	[4.45, 4.94]	-.51	-.76
Understanding	Mean of 7 component measures, standardized	Standardized Higher=Greater Understanding	[-.07, .09]	-.53	-.52
User evaluation of the communication	Mean of 5 items on the dimensions helpful, accurate, useful, credible, liking	Scale from 1-7 Higher=Greater Positive Evaluation	[4.25, 4.54]	-.17	-.68
Worry about exposure to UXO	Mean of 9 items about work at construction site, live near it, children play near it	Scale from 1-7 Higher=Greater Worry	[5.36, 5.63]	-1.10	1.08
Support for Construction	Mean of 6 items asking whether would support, work, or live near construction site	Scale from 1-7 Higher=Greater Support	[2.32, 2.61]	1.13	.88
Support for Increased Pay and Education	Mean of 6 items about increased pay and education for working at construction site	Scale from 1-7 Higher=Greater Support	[5.90, 6.10]	-1.44	3.65

Table II

Means of the outcome variables for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
Perceived likelihood [2]	.12	-.12	.001
Experienced fear (0-8 scale)	5.08	4.32	.002
Understanding [2]	-.07	.09	.009
User evaluation of the materials (1-7 scale)	4.34	4.44	.60
Worry about exposure to UXO (1-7 scale)	5.58	5.41	.18
Decision making about land redevelopment			
Support for Construction (1-7 scale)	2.35	2.59	.07
Support for Increased Pay and Education (1-7)	6.01	5.99	.61

[1] The provided p-values are for the main effect of Graph Format from the full ANCOVA.

[2] This measure is in standardized form, being an aggregate of variables on different scales.

Table III

Correlations among the outcome measures

Variable	Experienced Fear	Understanding	User Evaluation of the Communication	Worry	Support for Construction Activities	Support for Increased Pay and Education
Perceived Likelihood	.32***	-.32***	-.16**	.52***	-.35***	.21***
Experienced Fear		-.06	.09	.46***	-.32***	.23***
Understanding			.16**	-.10†	.09	.13*
User Evaluation of the Communication Materials				-.03	.09	.17**
Worry					-.67***	.40***
Support for Construction Activities						-.29***

† p < .10, * p < .05; ** p < .01; *** p < .001

Figure Captions

Figure 1. Hypothesized model of how graph format influences decision making. According to this model, graph format influences size of the perceived likelihood, experienced fear, understanding of the risk magnitudes, and the user's evaluation of the risk communication, all of which influence worry, which in turn influences decision making.

Figure 2. Risk communication information, as shown in two of the four conditions: (a) foreground-only graph format, actual probabilities condition; (b) foreground and background graph format, smaller probabilities condition.

Figure 3. A structural equations model of how graph format influences support for construction activities and support for increased pay and education for construction workers. Coefficients are standardized beta weights.

Figure 1

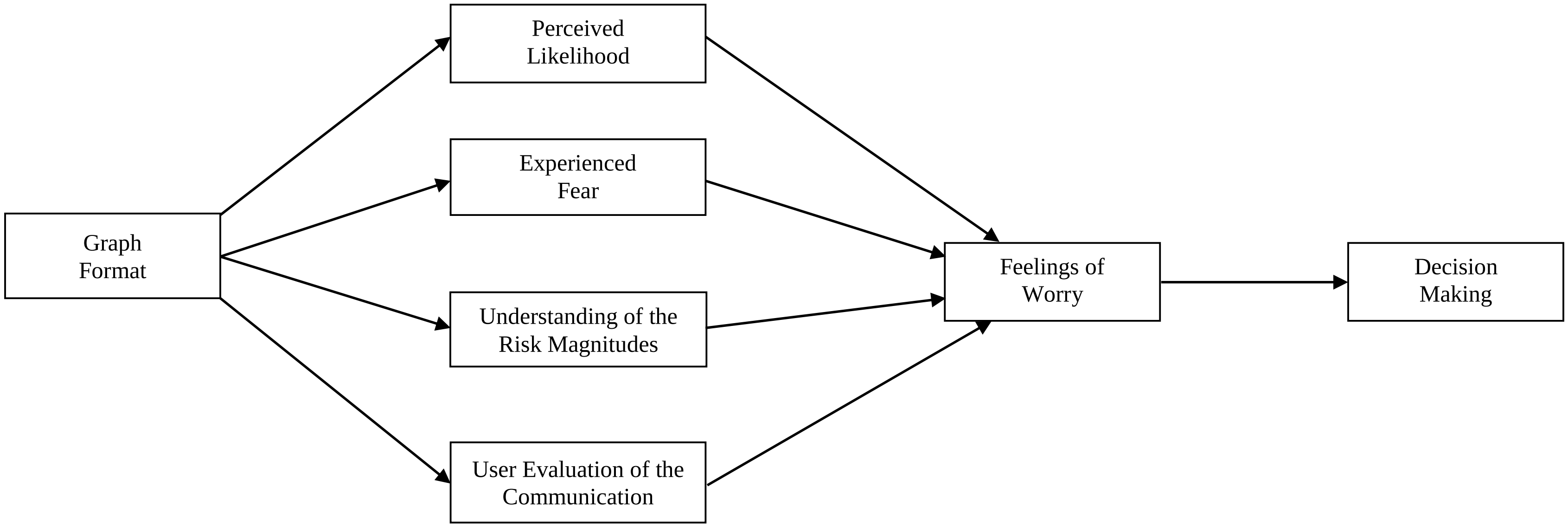

Graph
Format
Perceived
Likelihood
Experienced
Fear
Understanding of the
Risk Magnitudes
User Evaluation of the
Communication
Feelings of
Worry
Decision
Making

Figure 2a

Fort Ord, California was an Army base from 1917 until 1994. Soldiers practiced firing ammunition in the fields there. They fired anti-tank projectiles, grenades, landmines, mortars, rockets, and other kinds of ammunition. This ammunition did not always explode when fired. Some of it may still be left in the ground. Any contact with it could cause an explosion.

Now, there is a new plan for 310 acres of the land. Monterey Peninsula College wants to construct buildings, roads, and parking lots there. Construction workers run the risk of coming across unexploded ammunition. They can be hurt or killed if it goes off.

Imagine you live in a community near Fort Ord. You need to decide if the risk is small enough to allow construction. The risk of workers hitting ammunition is larger if they dig deeper. For example, buildings require more digging than do parking lots.

The graph below displays the chance that an individual worker will be exposed to unexploded ammunition at each of 4 digging depths.

Number of workers (per 100) exposed to unexploded ammunition when digging

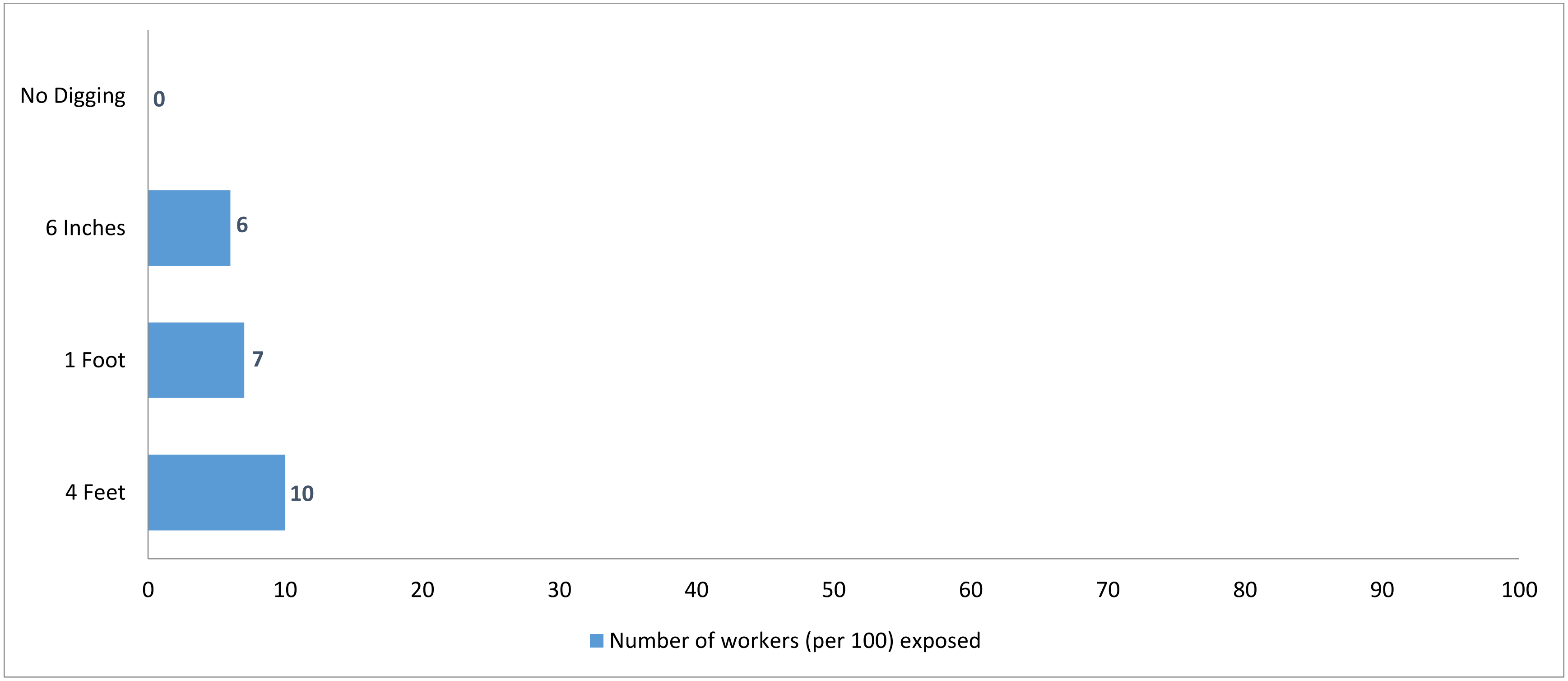


Figure 2b

[Same text as in Figure 2a.]

Number of workers (per 1000) exposed to unexploded ammunition when digging

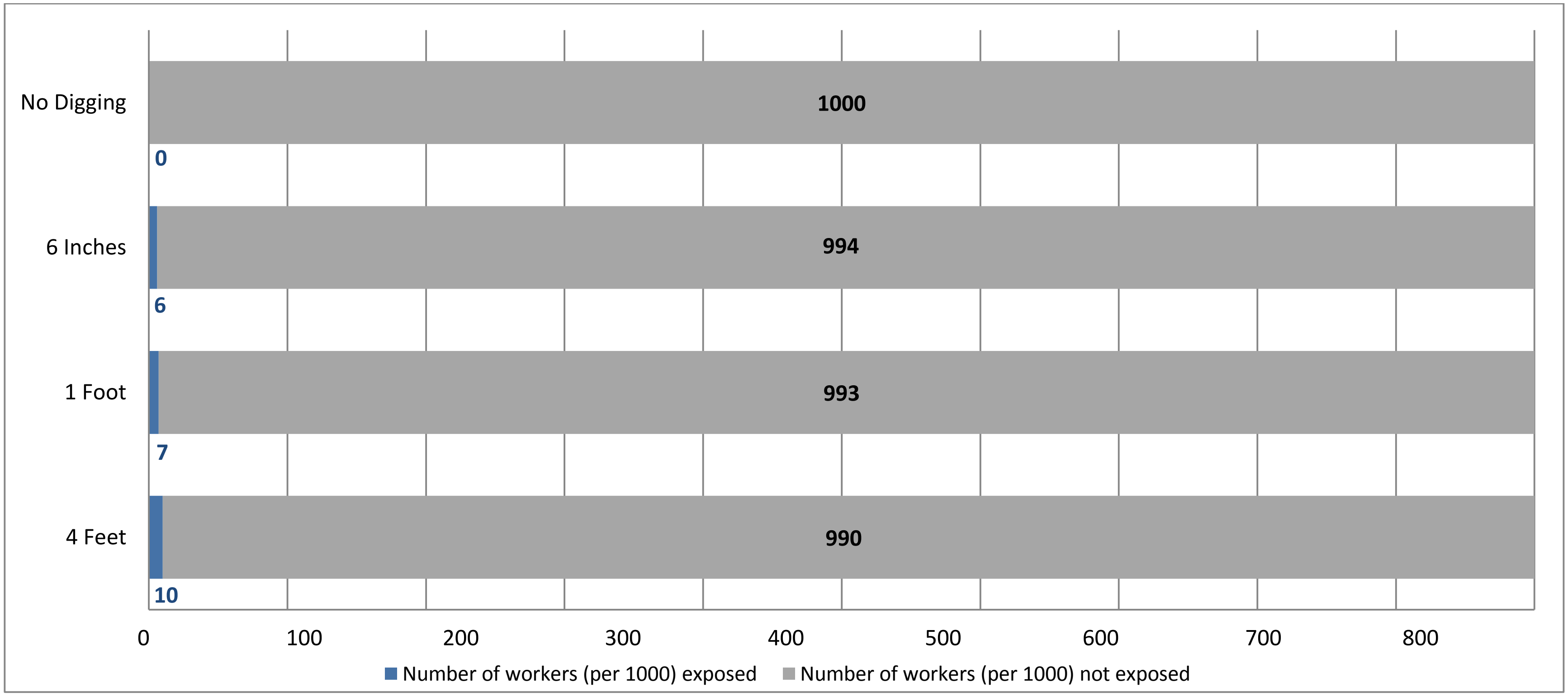
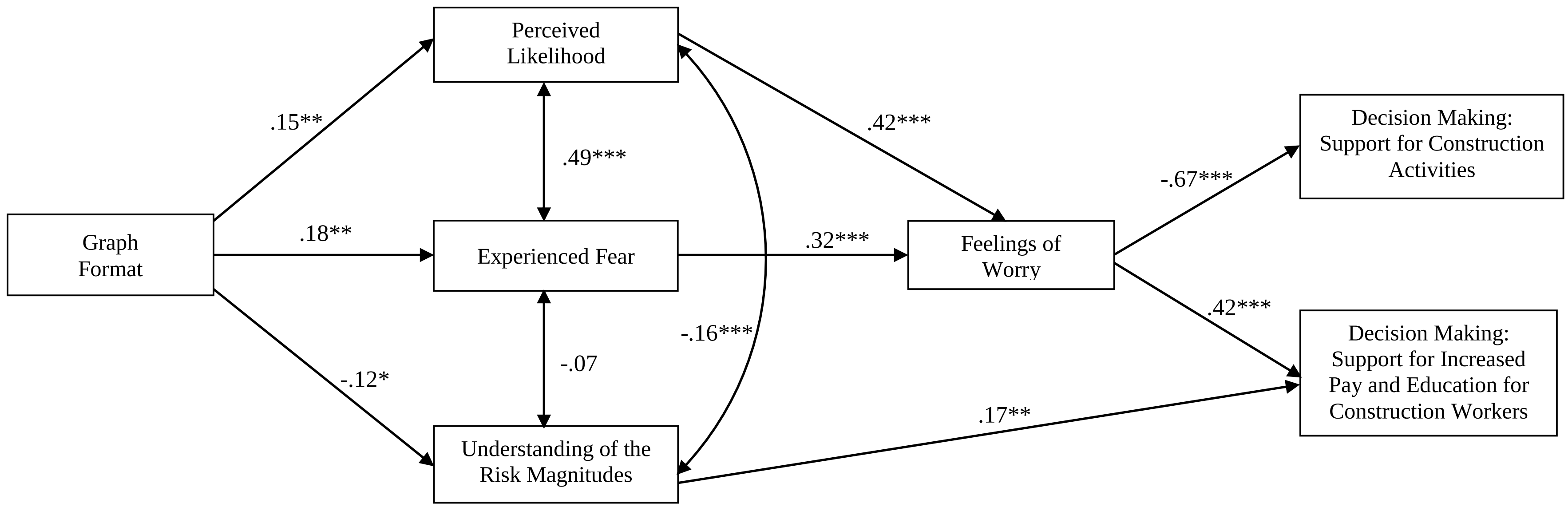


Figure 3



†p < .10, *p < .05, **p < .01, ***p < .001

Supplementary Materials for “Designing Graphs to Communicate Risks: Understanding How the Choice of Graphical Format Influences Decision Making”

Goals of the Supplementary Material

We measured each of our constructs with multiple measures. For exposition and space reasons, we combined them into one overall measure of each construct for the main paper. However, there are theoretical reasons to believe that different measures of each construct (e.g., understanding) may produce different results (e.g., Stone et al., 2015). Here, we provide more detail on the construction and scoring of each of the component measures and how they were influenced by the manipulations. We organize the Supplementary Material according to the construct under investigation: a) perceived likelihood, b) experienced fear, c) understanding, d) user evaluation of the communication materials, and e) worry. We do not discuss decision making here, since that variable was already broken down in the main paper.

Perceived Likelihood

Perceived likelihood was comprised of six component measures.

Component measures

1. <u>Overall chance of exposure</u>

This measure was based on Likert-type scales used by Shepperd et al. (2013), Stone et al. (2003), and others. Participants were asked one question about their overall risk of exposure: "What do you think is the overall chance of being exposed to unexploded ammunition when working at this building site?" (from 1=extremely low to 7=extremely high)

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [3.74, 4.13], skewness = 0.00, kurtosis = -0.98.

2. <u>Chance of exposure if a construction worker</u>

This measure was based on Likert-type scales used by Shepperd et al. (2013), Stone et al. (2003), and others. Participants were asked one question about the risk of exposure to unexploded ammunition if one was a construction worker: "Imagine you are a **construction worker** at this building site. What do you think is your chance of **being exposed** to unexploded ammunition during the time you work there?" (from 1=no chance to 7=certain to happen)

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [3.87, 4.23], skewness = 0.04, kurtosis = -1.12.

3. <u>Chance of injury if a construction worker</u>

This measure was based on Likert-type scales used by Shepperd et al. (2013), Stone et al. (2003), and others. Participants were asked one question about the risk of injury if one was a construction worker: "Imagine you are a **construction worker** at this building site. What do you think is your chance of **being injured** by unexploded ammunition during the time you work there?" (from 1=no chance to 7=certain to happen)

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [3.62, 3.97], skewness = 0.23, kurtosis = -1.00.

4. <u>Chance of exposure if a community member</u>

This measure was based on Likert-type scales used by Shepperd et al. (2013), Stone et al. (2003), and others. Participants were asked one question about the risk of exposure to unexploded ammunition if one was a community member: "Now imagine you are a **community member** of the local community of Fort Ord. What do you think is your chance of **being exposed** to unexploded ammunition around this building site?" (from 1=no chance to 7=certain to happen)

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [2.61, 2.94], skewness = 0.72, kurtosis = 0.09.

5. Chance of injury if a community member

This measure was based on Likert-type scales used by Shepperd et al. (2013), Stone et al. (2003), and others. Participants were asked one question about the risk of injury if one was a community member: “Imagine you are a **community member** of the local community of Fort Ord. What do you think is your chance of **being injured** by unexploded ammunition around this building site?” (from 1=no chance to 7=certain to happen)

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [2.23, 2.51], skewness = 1.00, kurtosis = 0.71.

6. Size of “out of 450” estimates

This measure was based on an approach taken by Chua et al. (2006) and Stone et al. (2015) to ask participants to recall the risk statistics and measure perceived likelihood by seeing how large these recollections are. Specifically, we asked participants to state how many out of 450 construction workers would be exposed to unexploded ammunition at each of the four digging depths, for example, “Six inches of Digging: _______ out of 450 construction workers would be exposed to unexploded ammunition at this site.”

Scoring. We averaged the four estimates; these averages had a considerable amount of skew (skewness = 2.96) and kurtosis (kurtosis=10.58), so we conducted a logarithmic transformation on the averaged values.

Interpretation. The range of scores was 0 to 2.61 (0 to 406.25 prior to the transformation), with higher scores indicating greater perceived likelihood.

Descriptive statistics. The 95% confidence interval for the mean was [1.13, 1.26], skewness = 0.19, kurtosis = -0.76.

Results

In keeping with the results in the main body of the paper, we z-scored each of the component measures and averaged them to comprise an overall measure of perceived likelihood. We conducted 2 (Graph Format) by 2 (Probability Level) ANCOVAs (with the same covariates as described in the main text) on the summary measure of perceived likelihood as well as on each of the component measures. Note the results with overall perceived likelihood are the same as those presented in the main text. The results for the main effect of Graph Format are presented in Table S1. There were no significant interactions with Probability Level (all interaction p's $> .05$).

As is evident in Table S1, the foreground-only graphical display led to a greater perception of how large the risks were on each of the component measures. All of these were significant with the exception of the chance of injury to a community member, which was marginally significant at $p = .06$.

Table S1

Mean levels of perceived likelihood for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
Perceived Likelihood [2]	.12	-.12	.001
Overall chance of exposure (1-7 scale)	4.20	3.68	.002
Chance of exposure:construction worker (1-7 scale)	4.25	3.85	.006
Chance of injury:a construction worker (1-7 scale)	3.95	3.64	.05
Chance of exposure:community member (1-7 scale)	2.91	2.64	.02
Chance of injury:community member (1-7 scale)	2.46	2.28	.06
Size of "out of 450" estimates (in log units)	1.29	1.10	< .001

[1] The provided *p*-values are for the main effect of Graph Format from the full ANCOVA.

[2] This measure is in standardized form.

Experienced Fear

This measure was adapted from the work of Lerner et al. (2003). Participants were told "Assuming you are a construction worker living near this building site, think about how you felt when you read the information about unexploded ammunition." Participants then responded to four items that assessed their fear, as well as eight other items not reported here.

Component measures

1. Worried

Participants responded to the statement "I felt worried when I read the information about unexploded ammunition" by circling a number from 0=Did not feel the emotion the slightest bit to 8=Felt the emotion more than ever before.

Scoring. No transformation necessary.

Interpretation. The range of scores was 0-8, with higher scores indicating more experienced fear.

Descriptive statistics. The 95% confidence interval for the mean was [4.91, 5.38], skewness = -.79, kurtosis = -.12.

2. Fearful

Participants responded to the statement "I felt fearful when I read the information about unexploded ammunition." by circling a number from 0=Did not feel the emotion the slightest bit to 8=Felt the emotion more than ever before.

Scoring. No transformation necessary.

Interpretation. The range of scores was 0-8, with higher scores indicating more experienced fear.

Descriptive statistics. The 95% confidence interval for the mean was [4.94, 5.45], skewness = -0.75, kurtosis = -0.31.

3. Frightened

Participants responded to the statement “I felt frightened when I read the information about unexploded ammunition” by circling a number from 0=Did not feel the emotion the slightest bit to 8=Felt the emotion more than ever before.

Scoring. No transformation necessary.

Interpretation. The range of scores was 0-8, with higher scores indicating more experienced fear.

Descriptive statistics. The 95% confidence interval for the mean was [4.47, 5.02], skewness = -0.55, kurtosis = -0.83.

4. Terrified

Participants responded to the statement “I felt terrified when I read the information about unexploded ammunition” by circling a number from 0=Did not feel the emotion the slightest bit to 8=Felt the emotion more than ever before.

Scoring. No transformation necessary.

Interpretation. The range of scores was 0-8, with higher scores indicating more experienced fear.

Descriptive statistics. The 95% confidence interval for the mean was [3.40, 3.98], skewness = -0.04, kurtosis = -1.28.

Results

In keeping with the results in the main body of the paper, we averaged the four components (i.e., the 4 fear items) to comprise an overall measure of experienced fear. We conducted 2 (Graph Format) by 2 (Probability Level) ANCOVAs (with the same covariates as described in the main text) on the summary measure of experienced fear as well as on the individual items. Note the results with experienced fear are the same as those presented in the main text. The results for the main effect of Graph Format are presented in Table S2. There were no significant interactions with Probability Level (all interaction p's > .22).

As is evident in Table S2, the foreground-only graphical display led to greater fear than the foreground and background graphical display on each of the four fear items.

Table S2

Mean levels of experienced fear for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
Experienced Fear (0-8 scale)	5.08	4.32	.002
Worried (0-8 scale)	5.49	4.81	.002
Fearful (0-8 scale)	5.61	4.79	.001
Frightened (0-8 scale)	5.11	4.38	.009
Terrified (0-8 scale)	4.10	3.29	.005

[1] The provided *p*-values are for the main effect of Graph Format from the full ANCOVA.

Understanding

We assessed understanding using seven different measures adapted from previous research (Bruine de Bruin et al., 2014; Chua et al., 2006; Cuite et al., 2008; Miron-Shatz et al., 2009; Stone et al., 2015; Weinstein, 1999; Wright et al., 2009). These included measures of understanding of both absolute and relative risk magnitudes (Stone et al., 2015; Weinstein, 1999), as well as a measure of the participant's subjective evaluation of their understanding, and were chosen to assess participants' understanding of both the gist of the information as well as the verbatim information (Hawley et al., 2008; Reyna, 2008; Reyna & Brainerd, 2008; Stone et al., 2015; Zikmund-Fisher, 2013). Specifically, two of the measures assessed understanding of absolute risk magnitudes, four understanding of relative risk magnitudes, and one perceived understanding.

Component measures

1. Absolute correctness of recall items (understanding of absolute risk magnitudes)

This measure was based on work by Stone et al. (2015) and others who asked participants to recall the information provided to them. Specifically, we asked participants four fill-in-the-blank questions pertaining to each of the digging depths, where participants responded to, for example, "I think that ___ out of ____ workers at this building site will be exposed to unexploded ammunition when digging one foot into the ground."

Scoring. For each of the four digging depths, we calculated the estimated percentage chance of exposure to unexploded ammunition. Conceptually, we are interested in how close these estimations are to the actual percentages at each of the digging depths.

One approach would be to calculate the absolute differences at each of the digging depths and sum the absolute differences. This approach is problematic, however, given that some participants substantially overestimated the risks, producing extreme skew (skewness = 5.77) and kurtosis (kurtosis=53.88). (See Miron-Shatz, Hanoch, Graef, & Sagi, 2009, for a discussion of this issue.) Thus, consistent with the approach taken by Stone et al. (2015), for each digging depth we scored correctness on a 3-point scale (correct, close to correct, far from correct) and summed these four values to get an overall score for this measure.

For example, when scoring the percentage chance of exposure when digging 6 inches for the actual probabilities condition, the correct answer was 6% and thus received a score of two. Perceived percentages between 3% and 10% (inclusive, but not including 6%) received a score of one, and responses outside this range received a score of zero. The ranges for the “close” (1-point) responses were chosen with the goal of having roughly as many “close” and “far” responses and to be as symmetric as possible around the correct answer, although given the skew in the distributions this was not always possible, and we also looked for natural breaking points in the frequency distributions.

Interpretation. The range of scores was 0-8 (0-2 for each digging depth), with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [5.19, 5.78], skewness = -0.54, kurtosis = -1.10.

2. <u>Absolute correctness of “out of 450” items</u> (understanding of absolute risk magnitudes)

This measure was based on an approach taken by Stone et al. (2015) to use a denominator different from that used in the communication when assessing recall to better capture participants’ gist by making it more difficult to respond with the verbatim information. Specifically, we took the responses to the questions where we asked participants to state how many "out of 450 construction workers" would be exposed to unexploded ammunition at each of the digging depths (Perceived Likelihood Component measure #6) and scored how close they were to the true values.

Scoring. As again just summing the absolute deviations produced considerable skew (skewness = 3.22) and kurtosis (kurtosis = 12.27), we scored correctness on a 3-point scale (correct, close to correct, far from correct) at each of the digging depths and summed these four values to get an overall score for this measure.

Interpretation. The range of scores was 0-8 (0-2 for each digging depth), with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [5.14, 5.70], skewness = -0.54, kurtosis = -0.99.

3. <u>Correlation of the recall estimates with the actual percentages</u> (understanding of relative risk magnitudes)

Consistent with one of the measures of understanding of relative risk magnitudes used in Stone et al. (2015), we took the four percentage estimates from (1) above and correlated these estimates with the actual percentage chances of exposure. Thus, whereas (1) examined how accurate one's overall estimates were across the digging depths, this measure examined the relative risk rankings for the four digging depths.

Scoring. We calculated the correlation between the estimated percentage chances of exposure with the actual percentage chances. Most participants' responses were highly correlated (95% were .77 or higher) with the actual likelihoods, which produced substantial negative skew (skewness = -6.87) and positive kurtosis (kurtosis = 65.01) in the correlations. Thus, we conducted a 5% winsorization on the left side of the distribution.

Interpretation. The range of scores was .77 (due to the winsorization) to 1, with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [.95, .96], skewness = -1.62, kurtosis = 1.82.

4. Correlation of "out of 450" estimates with actual percentages (understanding of relative risk magnitudes)

Also in keeping with the approach taken by Stone et al. (2015), we took the estimates from (2) above and correlated these estimates with the actual numbers.

Scoring. We calculated the correlation between the estimates and the actual numbers. As with (3), most participants' responses were highly correlated (95% were .79 or higher) with the actual numbers, which produced substantial negative skew (skewness = -7.28) and positive kurtosis (kurtosis = 59.12) in the correlations. Thus, we conducted a 5% winsorization on the left side of the distribution.

Interpretation. The range of scores was .79 (due to the winsorization) to 1, with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [.95, .96], skewness = -1.42, kurtosis = 1.12.

5. Tradeoff of risk with time items (understanding of relative risk magnitudes)

As discussed by Cuite et al. (2006), one way of assessing understanding is to determine whether participants can conduct "tradeoff" operations. Thus, we asked participants four questions regarding whether working at one digging depth for a certain amount of time would be more or less risky than working at a safer digging depth for a portion of that time and at a riskier digging depth for a different portion of time. An example item is: "Assume you work at the construction site for 2 months. Which would provide the **smallest chance** of exposure to unexploded ammunition? A. Digging at 6 inches for the first month and at 4 feet for the second month; B. Digging at 1 foot for both months."

Scoring. Each of the items was scored as correct or not and the four items were summed.

Interpretation. The range of scores was 0-4 (correct or not per item), with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [2.41, 2.64], skewness = 0.17, kurtosis = -0.80.

6. <u>Which increase in risk is greater items</u> (understanding of relative risk magnitudes)

Many researchers have assessed understanding by asking participants to rank a set of risks (e.g., Weinstein, 1999). As a simple ranking was too straightforward here (since it was obvious that increased digging depths would increase the risk) we asked two questions that required participants to understand not only the ordinal relationships, but also the extent to which they differed in risk. In particular, we asked participants two questions about which change in digging depth would produce the greatest increase in risk. An example item is: "For which of the following comparisons is the **increase in chance of exposure** greater? A. A change from no digging to six inches of digging. B. A change from six inches to one foot of digging."

Scoring. Each of the items was scored as correct or not and the two items were summed.

Interpretation. The range of scores was 0-2 (correct or not per item), with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [1.54, 1.67], skewness = -0.93, kurtosis = -0.20.

7. Subjective understanding

A number of researchers have assessed understanding by asking participants to rate their subjective understanding (e.g., Stone et al., 2015; Wright et al., 2009). We asked one question about subjective understanding of the risk magnitude information: "How well did you understand the information you were just given about how many people would be exposed to unexploded ammunition during construction on this building site?" (from 1=not at all to 7=very well).

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating greater understanding.

Descriptive statistics. The 95% confidence interval for the mean was [5.55, 5.88], skewness = -1.22, kurtosis = 0.90.

Results

In keeping with the results in the main body of the paper, we z-scored each of the component measures and averaged them to comprise an overall measure of understanding. Additionally, Stone et al. (2015) suggested that foreground-only graphical displays may be particularly poor for understanding of absolute risk magnitudes (see also Reyna, 2008). Thus we also constructed two additional summary understanding variables, one by averaging the (z-scored) measures of understanding of absolute risk magnitudes, and one by averaging the (z-scored) measures of understanding of relative risk magnitudes.

We conducted 2 (Graph Format) by 2 (Probability Level) ANCOVAs (with the same covariates as described in the main text) on each of the summary measures of understanding as well as on each of the component measures. Note the results with overall understanding are the same as those presented in the main text. The results for the main effect of Graph Format are presented in Table S3. There were no significant interactions with Probability Level (all interaction p's > .10).

As is evident in Table S3, descriptively the foreground-only graphical display produced less understanding then did the foreground and background graphical display for all seven component measures of understanding. However, the effects with the relative measures were smaller and not generally statistically significant. These results are consistent with those of Stone et al. (2015), who found that in comparison to a purely numerical display, a foreground-only graphical display reduced understanding, but primarily on measures of understanding of absolute risk magnitudes as opposed to relative risk magnitudes.

Table S3

Mean levels of understanding for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
Overall Understanding [2]	-.07	.09	.009
Understanding of Absolute Risk Magnitudes [2]	-.14	.13	.005
Correctness of recall items (0-8 scale)	5.23	5.74	.04
Correctness of "out of 450" items (0-8 scale)	5.05	5.77	.003
Understanding of Relative Risk Magnitudes [2]	-.08	.07	.08
Correlation of recall with actual percentages	.95	.96	.38
Correlation of "out of 450" with actual %'s	.95	.96	.15
Tradeoff of risk with time items (0-4 scale)	2.52	2.53	.58
Which risk increase is greater items (0-2 scale)	1.55	1.66	.07
Subjective Understanding (1-7 scale)	5.56	5.86	.06

[1] The provided p-values are for the main effect of Graph Format from the full ANCOVA.

[2] This measure is in standardized form.

.

User Evaluation of the Communication Materials

Five items were adapted from previous research assessing how positively participants evaluate the presentation of risk information (e.g., Bruine de Bruin et al., 2013; Dolan & Iadarola, 2008; Schapira et al., 2006). We reminded participants that they had been shown information about the chance of exposure at different digging levels and asked them to respond to that information in particular.

Component measures

1. <u>Helpful</u>

Participants were asked “How helpful did you find that information for decision making about the risks?” and circled a number from 1=Not helpful to 7=Very Helpful.

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating a more positive evaluation of the communication materials.

Descriptive statistics. The 95% confidence interval for the mean was [4.59, 4.97], skewness = -0.37, kurtosis = -0.81.

2. <u>Accurate</u>

Participants were asked “How accurate or true do you think the information was?” and circled a number from 1=Not Accurate to 7=Very accurate.

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating a more positive evaluation of the communication materials.

Descriptive statistics. The 95% confidence interval for the mean was [4.02, 4.35], skewness = -0.02, kurtosis = -0.64.

3. Useful

Participants were asked "How useful would you have found that information for making decisions about the risks?" and circled a number from 1=Not At All Useful to 7=Very Useful.

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating a more positive evaluation of the communication materials.

Descriptive statistics. The 95% confidence interval for the mean was [4.49, 4.88], skewness = -0.40, kurtosis = -0.77.

4. Liking

Participants were asked "How much did you like the way the information regarding the chance of exposure was presented to you?" and circled a number from 1=Did Not Like It to 7=Liked It a Lot.

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating a more positive evaluation of the communication materials.

Descriptive statistics. The 95% confidence interval for the mean was [4.11, 4.51], skewness = -0.32, kurtosis = -0.83.

5. Credible

Participants were asked "How credible did you find the information presented to you?" and circled a number from 1=Not at all credible to 7=Extremely credible.

Scoring. No transformation necessary.

Interpretation. The range of scores was 1-7, with higher scores indicating a more positive evaluation of the communication materials.

Descriptive statistics. The 95% confidence interval for the mean was [3.82, 4.17], skewness = -0.04, kurtosis = -0.75.

Results

In keeping with the results in the main body of the paper, we averaged the five component measures to comprise an overall measure of user evaluation of the communication materials. We conducted 2 (Graph Format) by 2 (Probability Level) ANCOVAs (with the same covariates as described in the main text) on the summary measure of user evaluation as well as on each of the component measures. Note the results with the overall measure of user evaluation are the same as those presented in the main text. The results for the main effect of Graph Format are presented in Table S4. There were no significant interactions with Probability Level (all interaction *p*'s > .07).

As is evident in Table S4, descriptively, the foreground and background graphical display was evaluated more highly than the foreground-only graphical display on each of the component measures. However, none of these effects approached significance.

Table S4

Mean levels of user evaluation for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
User evaluation of the communication mtrls (1-7 scale)	4.34	4.44	.60
Helpful (1-7 scale)	4.73	4.83	.65
Accurate (1-7 scale)	4.18	4.19	.98
Useful (1-7 scale)	4.66	4.71	.86
Liking (1-7 scale)	4.17	4.45	.24
Credible (1-7 scale)	3.97	4.01	.83

[1] The provided p-values are for the main effect of Graph Format from the full ANCOVA.

Worry about exposure to UXO

Nine items were adapted from previously used scales (Lipkus et al., 2005; Wright et al., 2009). These items assessed how worried, afraid, and anxious participants would be to work at the building site, live in the local area around Fort Ord, or let children play near the building site. Although the factor structure was well-explained by a one-factor solution, a 3-factor solution was readily interpretable as breaking down into those questions asking the participant about working at the building site, living near Fort Ord, and letting children play near the construction site. Here, we explore each of these individual factors as well as the overall measure of worry.

Component measures

1. Worry about working at the building site

Participants were asked three questions: (1) "If you were a construction worker at this building site, how **worried** would you be about finding unexploded ammunition during your work there?"; (2) "If you were a construction worker at this building site, how **afraid** would you be of finding unexploded ammunition during your work there?"; and (3) "If you were a construction worker at this building site, how **anxious** would you feel working at this building site?" All questions were on a 1-7 scale, with 1 = Not at all worried [afraid / anxious] and 7 = Extremely worried [afraid / anxious].

Scoring. The three responses were averaged into an overall measure of Worry about working at the building site.

Interpretation. The range of scores was 1-7, with higher scores indicating greater worry.

Descriptive statistics. The 95% confidence interval for the mean was [5.57, 5.87], skewness = -1.40, kurtosis = 1.84.

2. Worry about children near the building site

Participants were asked three questions: (1) "How **worried** would you be about letting children play near this building site?"; (2) "How **afraid** would you be about letting children play near this building site?"; and (3) How **anxious** would you feel if children were playing near this building site?" All questions were on a 1-7 scale, with 1 = Not at all worried [afraid / anxious] and 7 = Extremely worried [afraid / anxious].

Scoring. The three responses were averaged into an overall measure of Worry about children near the building site.

Interpretation. The range of scores was 1-7, with higher scores indicating greater worry.

Descriptive statistics. The 95% confidence interval for the mean was [6.17, 6.43], skewness = -2.18, kurtosis = 5.15.

3. Worry about living near to Fort Ord

Participants were asked three questions: (1) "How **worried** would you be about living in the local area around Fort Ord?"; (2) "How **afraid** would you be to live in the local area around Fort Ord?"; and (3) How **anxious** would you feel if you lived in the local area around Fort Ord?" All questions were on a 1-7 scale, with 1 = Not at all worried [afraid / anxious] and 7 = Extremely worried [afraid / anxious].

Scoring. The three responses were averaged into an overall measure of Worry about living near to Fort Ord.

Interpretation. The range of scores was 1-7, with higher scores indicating greater worry.

Descriptive statistics. The 95% confidence interval for the mean was [4.27, 4.66], skewness = -0.35, kurtosis = -0.83.

Results

In keeping with the results in the main body of the paper, we averaged the three components (i.e., the 9 items) to comprise an overall measure of worry. We conducted 2 (Graph Format) by 2 (Probability Level) ANCOVAs (with the same covariates as described in the main text) on the summary measure of worry as well as on each of the three component measures. Note the results with the overall measure of worry are the same as those presented in the main text. The results for the main effect of Graph Format are presented in Table S5. There were no significant interactions with Probability Level (all interaction *p*'s > .73).

As is evident in Table S5, despite the fact that the difference between the two display types did not reach significance for overall worry (as seen in the main text), the foreground-only graphical display did lead to greater worry than the foreground and background graphical display when only worry about working at the building site was considered, F (1, 288) = 6.96, p = .009. In retrospect, the fact that the display effect was strongest on this component makes sense, because the risk-magnitude information that was provided was in regards to digging at the site, not in terms of children playing there or living in the vicinity. Thus, it appears as if the lack of significance on overall worry occurred because the other types of worry (that weren't influenced by the display) diluted the effect on worry about working at the building site.

Table S5

Mean levels of worry for the foreground-only and foreground and background graphical conditions

Measure	Foreground Only	Foreground and Background	p-value [1]
Worry (1-7 scale)	5.58	5.41	.18
Worry about working at the building site (1-7 scale)	5.91	5.54	.009
Worry about children near the building site (1-7 scale)	6.32	6.27	.79
Worry about living near to Fort Ord (1-7 scale)	4.50	4.43	.57

[1] The provided p-values are for the main effect of Graph Format from the full ANCOVA.