

Do colored cells in risk matrices affect decision-making and risk perception? Insights from randomized controlled studies

Ruri Proto¹  | Gabriel Recchia¹ | Sarah Dryhurst^{1,2} | Alexandra L. J. Freeman¹

¹Winton Centre for Risk and Evidence Communication, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK

²Institute for Risk and Disaster Reduction, University College London, London, UK

Correspondence

Ruri Proto, Winton Centre for Risk and Evidence Communication, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WA, UK.
Email: rp718@cam.ac.uk

Funding information

Winton Centre for Risk & Evidence Communication

Abstract

Risk matrices communicate the likelihood and potential impact of risks and are often used to inform decision-making around risk mitigations. The merits and demerits of risk matrices in general have been discussed extensively, yet little attention has been paid to the potential influence of color in risk matrices on their users. We draw from fuzzy-trace theory and hypothesize that when color is present, individuals are likely to place greater value on reducing risks that cross color boundaries (i.e., the boundary-crossing effect), leading to sub-optimal decision making. In two randomized controlled studies, employing forced-choice and willingness-to-pay measures to investigate the boundary-crossing effect in two different color formats for risk matrices, we find preliminary evidence to support our hypotheses that color can influence decision making. The evidence also suggests that the boundary-crossing effect is only present in, or is stronger for, higher numeracy individuals. We therefore recommend that designers should consider avoiding color in risk matrices, particularly in situations where these are likely to be used by highly numerate individuals, if the communication goal is to inform in an unbiased way.

KEYWORDS

bias, color, decision making, fuzzy-trace theory, impact, likelihood, numeracy, risk communication, Risk matrix

1 | INTRODUCTION

Risk matrices illustrate the gravity of certain hazards as a function of their likelihood and potential impact and are used across the public and private sectors to make decisions about risk mitigation and about how to allocate resources (Cox, 2008; Duijm, 2015). They can communicate quantitative or qualitative information about risks, depending on whether the likelihood and impact axes are both numeric, as in the case of quantitative risk matrices (e.g., 0.6% chance of an event resulting in £100k cost increase), non-numeric categories, as in the case of qualitative risk matrices (e.g., unlikely to happen with a catastrophic impact), or a combination of the two, as in the case of semiquantitative risk matrices (e.g., 0.6% chance of an event happening with a catastrophic impact).

Risk matrices are widely used to communicate risks and support decision-making around risk mitigation. For exam-

ple, risk matrices have recently been used to assist with the development of therapeutic guidelines that mitigate health risks (Lemmens et al., 2022), communicate the threat of plastic pollution to aquatic wildlife across different species (Roman et al., 2022), visualize various weather- and non-weather-related risks to cereals production (Huet et al., 2022), and communicate risks of various natural hazards such as rock bursts (Kadkhodaei & Ghasemi, 2022), floods, and droughts (Cotti et al., 2022). Risk matrices are also used to support international and national risk management practices, e.g., those of the U.K. Cabinet Office (2017), World Economic Forum (2021), and various international standards such as ISO 31010:2019 (International Organization for Standardization, 2019).

Commonly, a risk matrix will delineate different categories of risks through color codes that categorize the risk depending on its impact and likelihood levels, with red typically

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Risk Analysis* published by Wiley Periodicals LLC on behalf of Society for Risk Analysis.

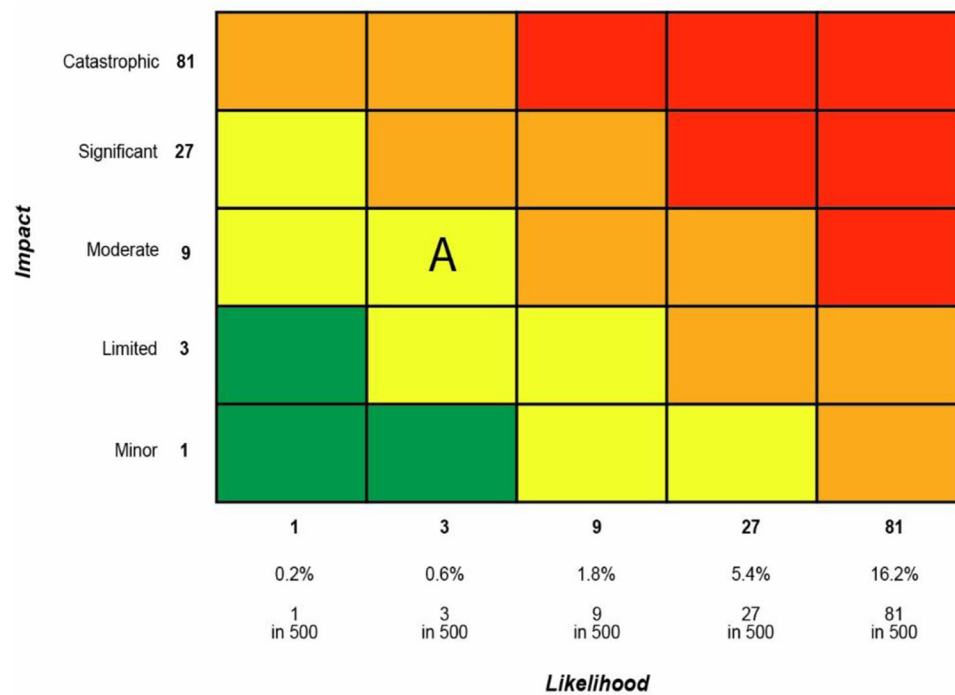


FIGURE 1 Example risk matrix format using color categories (cell shading format)

attributed to the most grave, followed by orange, yellow, and green for less grave risks (Levine, 2012). As an example, consider the semiquantitative matrix format in Figure 1, where the likelihood axis presents geometrically increasing probabilities (e.g., 0.3%, 1.8%, 5.4%), the impact axis presents ordinal categories of severity (e.g., significant, catastrophic), and each cell is assigned a color, meaning that Risk A in this case would be considered a “yellow” level risk. In this study, we employ a format that includes geometrically increasing probabilities, as this is a common nonlinear scaling used in practice. Additionally, we include geometric labels on both axes (i.e., 1, 3, 9, 27, 81), as previous research has found these labels enhance comprehension of the nonlinear nature of the axes (Sutherland et al., 2021).

There are extensive critiques that focus on the merits and demerits of quantitative risk matrices (Cox, 2008; Duijm, 2015; Levine, 2012), qualitative risk matrices (Cox et al., 2005), and emerging research critiquing semiquantitative risk matrices (Monat & Doremus, 2020; Sutherland et al., 2021) in general. Additionally, there is some evidence that presenting information about the likelihood and impact of an event along with a warning color in a risk matrix format helps individuals make objectively better decisions than presenting only the warning color with no information about likelihood and impact (Mu et al., 2018). However, while this study shows that the risk matrix format has advantages over a color-only format due to the additional information it presents, there is no consideration of the potential psychological effects of adding colors to otherwise blank risk matrices.

In this study, we draw from fuzzy-trace theory (Reyna & Brainerd, 1995) and investigate empirically whether adding

color-based, qualitative boundaries between cells in a matrix might motivate decision-makers to prefer mitigations that move a risk into a different color category, even if this change reduces risk less than an alternative mitigation of equal cost (Experiment 1). We also test whether decision-makers are likely to value risk reductions that cross color boundaries more than those that do not cross color boundaries despite the fact that both achieve the same absolute reduction in likelihood (Experiment 2).

Fuzzy-trace theory argues that when processing risk information, individuals use different types of cognitive representations that vary in precision (Reyna & Brainerd, 1995). The theory distinguishes between verbatim representations, which encode precise, detailed information (e.g., the numeric probability of an event happening), and gist representations, which encode surface-level, simpler, categorical (e.g., some risk vs. no risk) or ordinal (e.g., low risk vs. high risk) information. A key precept of fuzzy-trace theory is that although all these representations are encoded in parallel, individuals prefer to rely on the simplest information that clearly differentiates between options when making decisions (Reyna & Brust-Renck, 2020).

As matrix cell colors convey a qualitative, ordinal description of risk severity, they are likely to facilitate decision-making that relies on ordinal gist representations, rather than verbatim representations. For example, consider a case where someone has to choose between reducing either of two risks as in Figure 2. According to the colors, the decision would be reduced to a comparison of ordinal gist representations, that is, choosing between reducing risk A from “bad” to “good” (i.e., from A1: yellow to A2: green) and reducing risk B

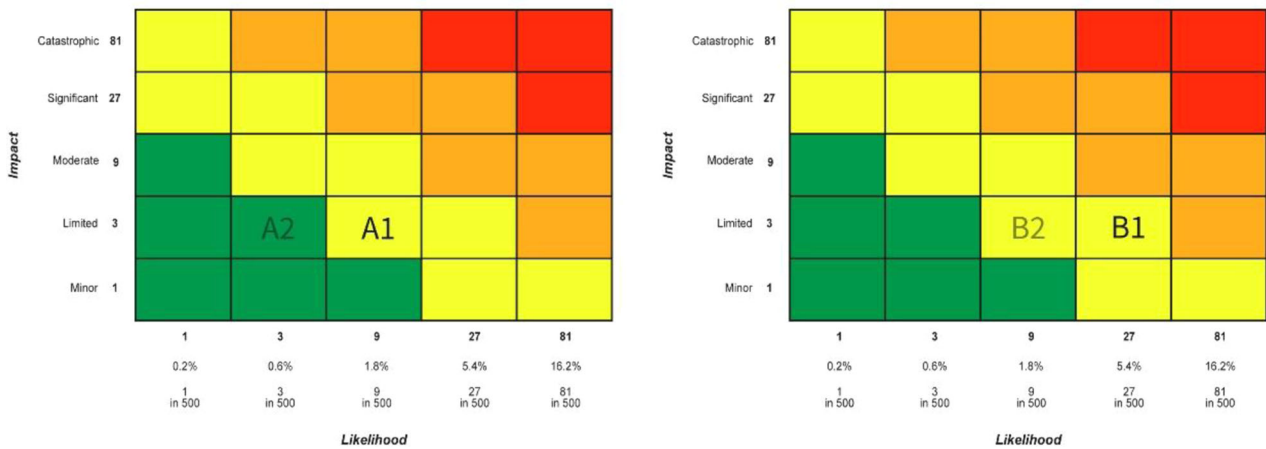


FIGURE 2 Risk reduction options in a cell shading format risk matrix: A1 to A2 is a mitigation that causes risk A to cross a color boundary; B1 to B2 is a mitigation that does not cause risk B to cross a color boundary.

from “bad” to “bad” (i.e., from B1: yellow to B2: yellow). As this is the simplest representation available, fuzzy-trace theory predicts that a decision maker will opt to reduce risk A, which seems superior to reducing risk B according to the simple color indications. However, note that reducing risk B from position B1 to B2 achieves a greater absolute reduction ($5.4\% - 1.8\% = 3.6\%$) in likelihood than reducing Risk A from A1 to A2 ($1.8\% - 0.6\% = 1.2\%$). This verbatim information is much more precise and complex than the color information, meaning it is less likely to be used in decision making.

Thus, individuals might prefer a risk reduction that crosses color boundaries even if another risk reduction that does not cross color boundaries achieves a greater actual reduction in the likelihood of the event, which would be evident when inspecting axis labels. We refer to this phenomenon as the boundary-crossing effect. By contrast, a risk matrix without color would provide fewer cues for qualitative distinctions, thus prompting participants to rely on verbatim representations, which might lead to more rational decision-making when reducing risks (i.e., a preference for those risk reductions that minimize the likelihood or impact of an event).

Considering the popularity of using colors in matrices, we sought a variation that might alleviate any biasing effects and developed what we call the color banding format (see Figure 3). Crossing color boundaries in this format might still be tempting because an ordinal gist distinction in risk severity is still suggested by the different colors. However, because the design adds a band of cells that are “half one color, half another” between each color category, a more continuous transition between risk categories is enforced. Mitigations moving a risk to a single cell on either dimension can no longer move it from one single color level directly to another, which might reduce the perceived degree of difference between levels, thus lessening participants’ preference for crossing color boundaries.

Arguably, the color banding format can be adapted to include different color split ratios (e.g., 90% one color—10% another color, 75% one color—25% another color). In this study, we restrict our matrix design to a color banding format wherein cells can either be “one color” or “half one color, half another” in order to minimize the possibility that the color split ratio could influence perceptions of what constitutes crossing a color boundary (e.g., crossing from a red cell to a another cell that is 90% red and 10% orange might not be perceived as “crossing a color boundary” in the same way in which crossing from a red cell to another cell that is 50% red and 50% orange is perceived).

Following the same argument—that participants would prefer risk reductions that cross color boundaries over those that do not—it is plausible that participants would also value risk reductions that cross color boundaries more than those that do not. In decision sciences, eliciting the amounts individuals are willing to pay for certain items or to make certain events happen is a widely employed approach to measure the subjective value they attribute to those items/events (He & Zhai, 2017). We carried out a pilot experiment (see Supporting Information Appendix A) prior to this study to determine a suitable method for eliciting participants’ willingness to pay to reduce a risk within the matrix, which we then used as a metric for assessing perceived value of hypothetical risk reductions.

1.1 | Objectives

In this study, we report two experiments. Experiment 1 was designed to test whether or not participants presented with risk matrices in either the cell shading format (Figure 1), color banding format (Figure 3), or a matrix with no color at all, would differ in their preferences for one potential risk reduction over another in a forced choice task. We pre-registered three hypotheses:

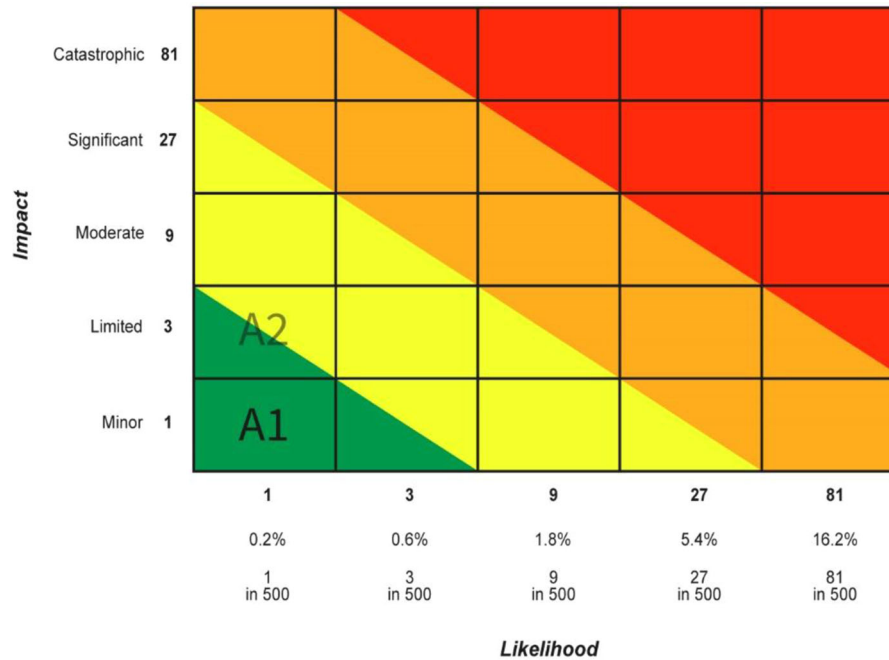


FIGURE 3 The color banding format, designed to make color groupings less categorical than the cell shading format

H1: Compared to participants presented with the no-color format, participants presented with the cell shading format will be more likely to prefer absolute risk reductions that cross color boundaries over other risk reductions that achieve a greater absolute risk reduction but do not cross color boundaries.

H2: Compared with participants presented with the no-color format, participants presented with the color banding format will be more likely to prefer absolute risk reductions that cross color boundaries over risk reductions that achieve a greater absolute risk reduction but do not cross color boundaries.

H3: Compared with participants presented with the cell shading format, participants presented with the color banding format will be less likely to prefer absolute risk reductions that cross color boundaries over other risk reductions that achieve a greater absolute risk reduction but do not cross color boundaries.

Experiment 2 was designed to test whether or not participants presented with risk matrices in the three possible formats showed differences in the values they assigned to different risk reduction options in a willingness-to-pay experimental format. We pre-registered four hypotheses:

H4a: In the cell shading condition, participants presented with risk matrices where the risk reduction crosses a color boundary will be willing to pay more for that reduction than participants presented with the same risk reductions but in the no-color matrix condition.

H4b: In the cell shading condition, participants presented with risk matrices where the risk reduction crosses a color boundary will be willing to pay more for that reduction than participants in the cell shading condition presented with the same risk reductions but in risk matrices modified such that these risk reductions do not cross a color boundary.

H4c: In the color banding condition, participants presented with risk matrices where the risk reduction crosses a color boundary will be willing to pay more for that reduction than participants presented with the same risk reductions but in the no-color matrix condition.

H4d: In the color banding condition, participants presented with risk matrices where the risk reduction crosses a color boundary will be willing to pay more for that reduction than participants in the color banding condition pre-

sented with the same risk reductions but in risk matrices modified such that this risk reductions do not cross a color boundary.

2 | EXPERIMENT 1: FORCED CHOICE

2.1 | Method

2.1.1 | Participants and procedure

Recruitment of UK-resident participants was facilitated by the ISO-accredited market research company Respondi. Interlocking quotas ensured that the sample was proportional to the UK population by gender and age, according to EURO-STAT 2019. Participants were invited to take part in the study and redirected to a Qualtrics survey, where they read the participant information sheet and digitally signed the consent form. Median completion time was 11 min, and participants were paid £2.97.

We recruited 5791 participants. Participants were excluded from analysis if they failed to complete the questionnaire in full (37.25% failure rate), if they failed the attention check (36.49% failure rate among remaining participants), or if they reported any kind of color blindness and were assigned to a color condition¹ (2.60% of remaining participants). The attention check (“How concerned should we be if you did not pay attention? To check that you are paying attention, please select 80 on the scale below:”) was answered using a moving slider toward the second part of the survey detailed in Experiment 2. Answers between 75 and 85 were coded as attention check passes to account for imprecision when moving the slider. The final analytic sample consisted of 2249 individuals (39% of the recruited sample).

Experiments 1 and 2 were part of the same survey, and our intended sample size was 3580 (see pre-registration here: <https://osf.io/bguxf>, and study materials and data here: <https://osf.io/9nvk3>). This sample size was intended to achieve 90% power for an independent-samples *t*-test in a set of four possible tests (corresponding to the four hypotheses in Experiment 2), with a conservatively corrected alpha level of 0.0125, and a small effect size (Cohen’s $d = 0.2$). The lower size of our analytic sample is mainly driven by the high attention check failure rate.

At the start of the survey, participants received instructions on how to read risk matrices (i.e., what the axes and geometric labels mean and explanations of the geometric scaling of risks on both likelihood and impact dimensions). Before being shown any of the experimental questions, participants completed questions assessing their personal experience with flooding, perceptions of the risk of flooding in the area where they live, and experience with risk matrices. Participants were then randomized to one of the three arms of Experiment 1.

Approximately 70% of participants had never used a risk matrix before, around 18% reported having used them rarely, around 8% reported using risk matrices sometimes, and fewer than 5% reported using risk matrices on a regular basis. With regard to flooding, around 80% of participants in each condition reported having no personal experience with flooding. Similarly, around 85% participants reported living in areas where they perceived the flooding risk to be low (see [Supporting Information Appendix B](#) for a detailed breakdown of covariate values across the whole sample).

At the very end of the survey (i.e., after completing Experiment 2 as well), participants completed a composite measure of numeracy consisting of the adaptive Berlin numeracy test (Cokely et al., 2012), the Schwartz numeracy scale (Schwartz et al., 1997), and an item from the expanded numeracy scale of Lipkus et al. (2001) in the light of previous research indicating that this combination of numeracy tests produces less-skewed distributions in public samples (e.g., Cokely et al., 2013; Sutherland et al., 2021). They were then asked to report their age, nationality, ethnicity, native language, household income, highest educational qualification, and political views in order to assess the degree to which our sample represented a diverse range of demographic characteristics (see [Supporting Information Appendix B](#)).

2.1.2 | Design

This experiment adopted a fully randomized between-subjects design. Participants were randomly allocated to one of the three risk matrix formats. Participants in the no-color condition represent the control group.

2.1.3 | Dependent variable: Risk comparison scores

Participant performance on four² risk reduction decision tasks was evaluated (one aimed at reducing impact, three aimed at reducing likelihood). For each task, participants had to choose between reducing two risks that were identical on one dimension (e.g., likelihood), but had different positions on the other dimension (e.g., impact) (see Figure 4). Risks were presented as hypothetical flood risks, referred to as “Risk A” and “Risk B.” Participants were informed that both risk reductions cost the same amount, but that they only had resources to reduce one of them, and were asked to state their choice. At the beginning of the survey, participants were explicitly instructed that the lighter text (i.e., “A2,” “B2”) represented the new risk positions after reduction, whereas the darker text (i.e., “A1,” “B1”) represented current risk positions, prior to any reduction.

¹ Results presented below also hold when participants who reported color blindness were excluded across all conditions.

² Participants were presented with five risk reduction decision tasks, but a survey error meant that one of the stimuli was not displayed correctly, and it was therefore removed from analysis.

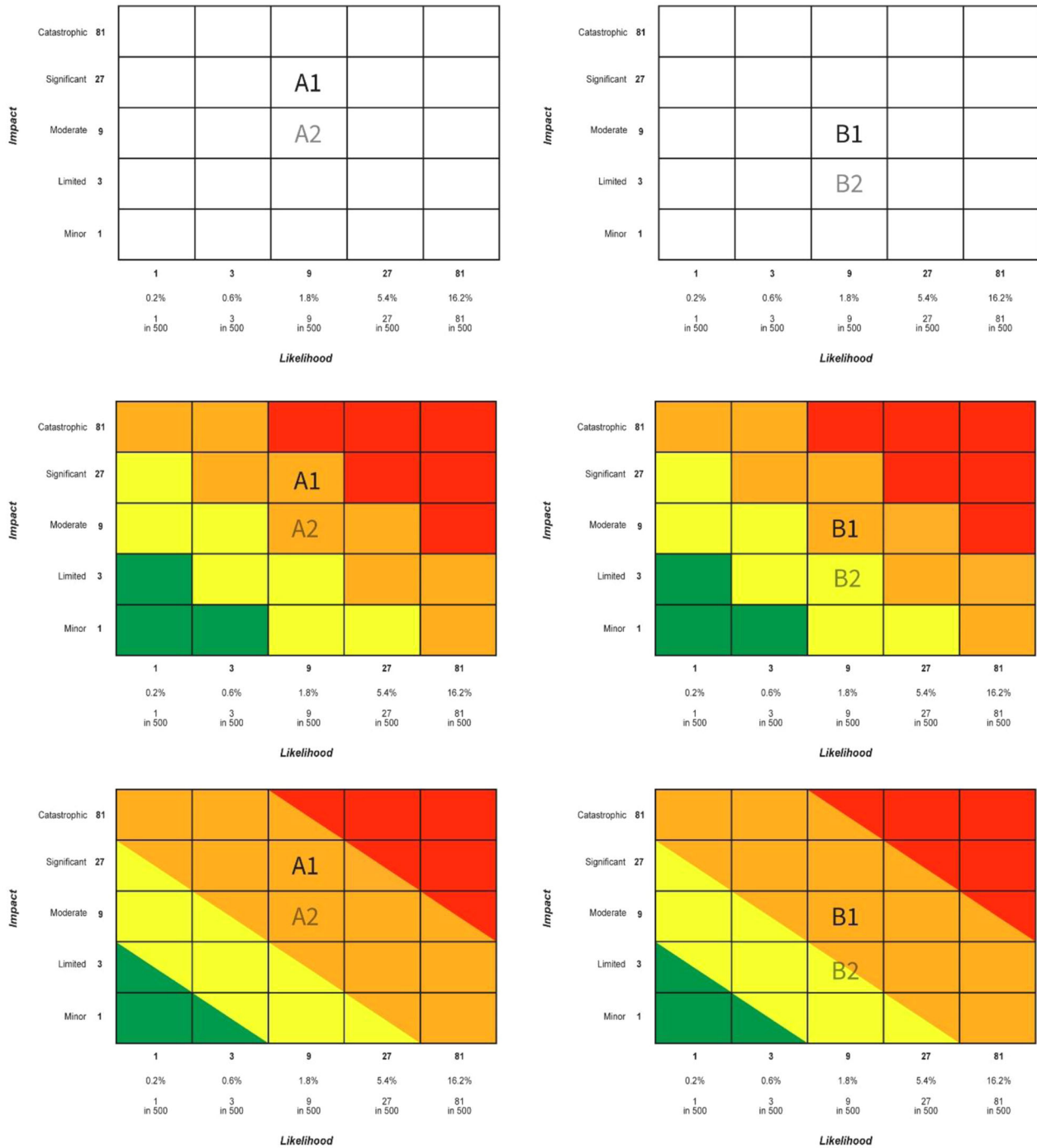


FIGURE 4 Example risk comparison stimuli from Experiment 1.1. Participants were shown one of the three pairs of matrices and asked whether they would prefer to reduce risk A (from position A1 to A2) or risk B (from position B1 to B2), if the costs of doing so were equal.

The exact placement of the risks across different impact and likelihood levels, and the absolute size of potential risk reductions differed across questions. Specifically, the amounts were selected such that the risk reduction option that crossed color boundaries always achieved a smaller absolute risk reduction than the alternative, which did not cross color boundaries. We anticipated that if the correct answer (the greater reduction) always corresponded to the choice that did not cross a color boundary, this might become obvious to

attentive participants. To break this pattern, and thus to prevent participants from becoming aware of our manipulation, two distractors were included in this set of questions (one aimed at reducing impact, one aimed at reducing likelihood). For these questions, the choice that gave the maximum absolute risk reduction was the one that crossed color boundaries, which is the complete opposite pattern to our task-relevant stimuli, and thus would prevent participants from picking up a generic pattern on which to base their answers, thereby

increasing the validity of our measure. The order in which the complete set of stimuli (task-relevant stimuli and distractors) was randomized across participants.

The risk comparison measure was constructed as follows. For each question where participants chose the option that achieved the greatest absolute risk reduction (i.e., the one that did not cross color boundaries), they were awarded one point. Points on all decision tasks (excluding distractors) were summed into an overall risk comparison score. Thus, a higher risk comparison score indicates more “rational” decision making, and, due to the nature of our stimuli, a preference for risk mitigations that do not cross color boundaries, as these were designed to achieve a greater absolute risk reduction than those that crossed color boundaries. The scale showed a slight ceiling effect but achieved a reasonable spread of scores with most participants scoring between 1 and 3 (range from 0 to 4) in each experimental group (see [Supporting Information Appendix A](#)). Compared to participants who failed the attention check, participants who passed it achieved higher risk comparison scores on average ($B = 0.215$, $p < 0.001$).

2.2 | Results

2.2.1 | Analytical approach

Variances in risk comparison scores were similar across experimental groups, as indicated by a nonsignificant Levene’s test ($F(2, 2246) = 0.674$, $p = 0.510$) and visual inspection of histograms. We chose covariates based on their whole-sample correlations with risk comparison scores and their variance explained in risk comparison scores when included in models alongside the independent variable (color format). This procedure is detailed in [Supporting Information Appendix C](#) and resulted in numeracy and flooding risk perceptions being selected as covariates. The assumption of independence from the independent variable held for both covariates.

Originally, we pre-registered a three-way between-subjects analysis of covariance (ANCOVA). However, we chose to fit a linear model that included the interaction between numeracy and condition, and flooding risk perceptions as a covariate, as this model improved variance explained (see [Supporting Information Appendix C](#)). The model including an interaction term between numeracy and format explained incremental variance in risk comparison scores over a model including only main effects for format, numeracy, and flooding risk perceptions ($F(4, 2238) = 4.375$, $p = 0.002$, $\eta^2_G = 0.004$). This model decreased AIC (Akaike Information Criterion) from 7769 to 7757, and decreased BIC (Bayesian Information Criterion) slightly from 7803 to 7802, indicating that the inclusion of the interaction term was justified when balanced against increased model complexity. This interaction suggests that the ANCOVA assumption of homogeneity or regression slopes is violated, thus justifying our deviation from the pre-registration.

2.2.2 | Findings

Participants in the color banding condition had lower risk comparison scores (i.e., performed worse in terms of maximizing risk reduction) than participants in the no-color and the cell shading conditions, with these effects increasing in strength as numeracy increased and reaching significance at average and high numeracy levels (see [Table 1](#)). There was no difference between the no-color and cell shading conditions across numeracy levels. As higher risk comparison scores indicate more rational, less biased decision-making, this pattern of results suggests H2 was supported, but H1 and H3 were not supported. The differences between formats at low, mean, and high numeracy levels, controlling for flooding risk perceptions, are also illustrated in [Figure 5](#).

3 | EXPERIMENT 2: WILLINGNESS TO PAY

As described above, this experiment was carried out within the same survey, with the same participants, as Experiment 1. The order in which participants completed experiments was randomized.

3.1 | Method

3.1.1 | Design

The experiment adopted a fully randomized design wherein participants were randomly allocated to one of five conditions: no color, cell shading with risk reductions that do not cross boundaries, cell shading with risk reductions that cross boundaries, color banding with risk reductions that do not cross boundaries, or color banding with risk reductions that cross boundaries (see [Figure 5](#)). Within each color format, the “crossing boundaries” group was compared against the “no crossing boundaries” group, and against the no-color group.

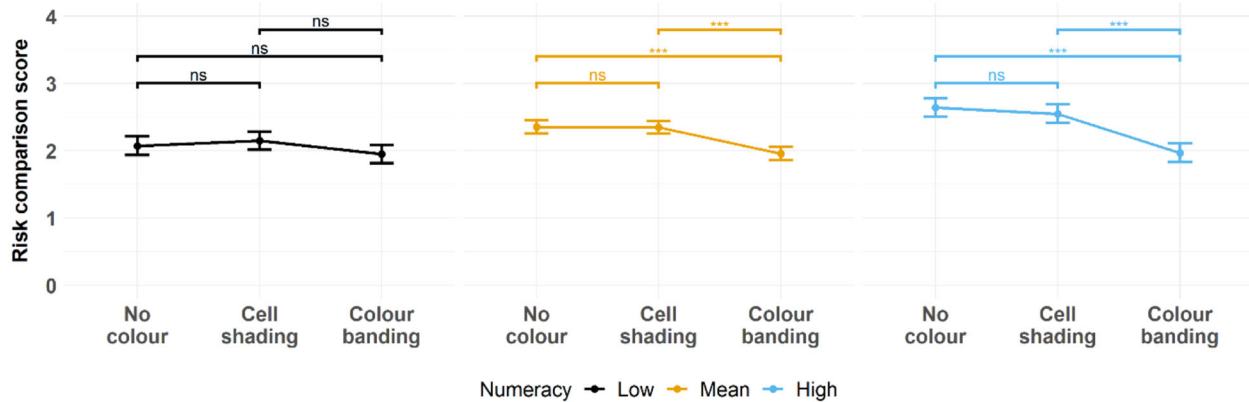
3.1.2 | Dependent variable: Willingness-to-pay amount

Participants were presented with three risk reduction options at low probability levels (reduction from 0.6% to 0.2%), medium probability levels (reduction from 1.8% to 0.6%), and high probability levels (reduction from 16.2% to 5.4%), corresponding to absolute reductions of 0.4%, 1.2%, and 10.8%, respectively. In each case, participants were asked to report how much they would be willing to pay out of a total budget of £100k (in £1k units) to achieve the risk reduction presented to them. The color pattern of the matrix was designed such that for some participants, the risk reductions crossed color boundaries (e.g., from red to orange), while for others, the same risk reductions did not cross color bound-

TABLE 1 Differences between formats at low, mean, and high numeracy levels, controlling for flooding risk perceptions

Reference level					
No color	Cell shading				
Numeracy	<i>B</i>	SE	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	0.075	0.010	0.758	0.449	0.505
Mean	–0.008	0.070	–0.109	0.913	0.914
High	–0.091	0.099	–0.919	0.358	0.460
No color	Color banding				
Numeracy	<i>B</i>	SE	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	–0.125	0.099	–1.260	0.208	0.311
Mean	–0.399	0.070	–5.682	<0.001	<0.001
High	–0.672	0.099	–6.800	<0.001	<0.001
Cell shading	Color banding				
Numeracy	<i>B</i>	SE	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	–0.201	0.098	–2.039	0.042	0.075
Mean	–0.391	0.070	–5.598	<0.001	<0.001
High	–0.582	0.100	–5.850	<0.001	<0.001

Note. *p* (B–H)—value adjusted with the Benjamini–Hochberg procedure. *B*—unstandardized regression weights. Low, mean, and high numeracy levels correspond to values one standard deviation below the mean of numeracy, mean numeracy, and one standard deviation above mean numeracy calculated across experimental groups. The overall model explained 4.3% of variance in risk comparison scores ($R^2 = 0.043$, adj- $R^2 = 0.040$).

**FIGURE 5** Marginal mean risk comparison scores and 95% confidence intervals in each format at different levels of numeracy, adjusted for flooding risk perceptions. Significance levels: ns— $p > 0.05$, ***— $p < 0.001$

aries (e.g., reduction from a red cell to another red cell) (see Figure 6). In order to make it possible for the same risk reduction to cross color boundaries for some participants, and not cross color boundaries for others, the proportions of colors in the matrix had to be altered accordingly. This is most evident in the color banding condition in Figure 5, where the proportion of orange relative to red changes depending on whether boundaries are crossed or not. Compared to participants who failed the attention check, participants who passed were willing to pay more to reduce risks on average ($B = 2.055$, $p = 0.003$).

3.1.3 | Manipulation check

To ensure our manipulation of probability levels was successful, we compared mean willingness-to-pay (WTP) amounts across conditions at within-subject level. Including probability level as a predictor reduced deviance compared to the unconditional model ($X^2(2) = 1405.1$, $p < 0.001$), with participants being willing to pay significantly less for risk reductions at low probability levels (i.e., small risk reductions) than at medium ($B = 5.534$, $t(4372.09) = 11.95$, $p < 0.001$) probability levels (medium reductions) and high

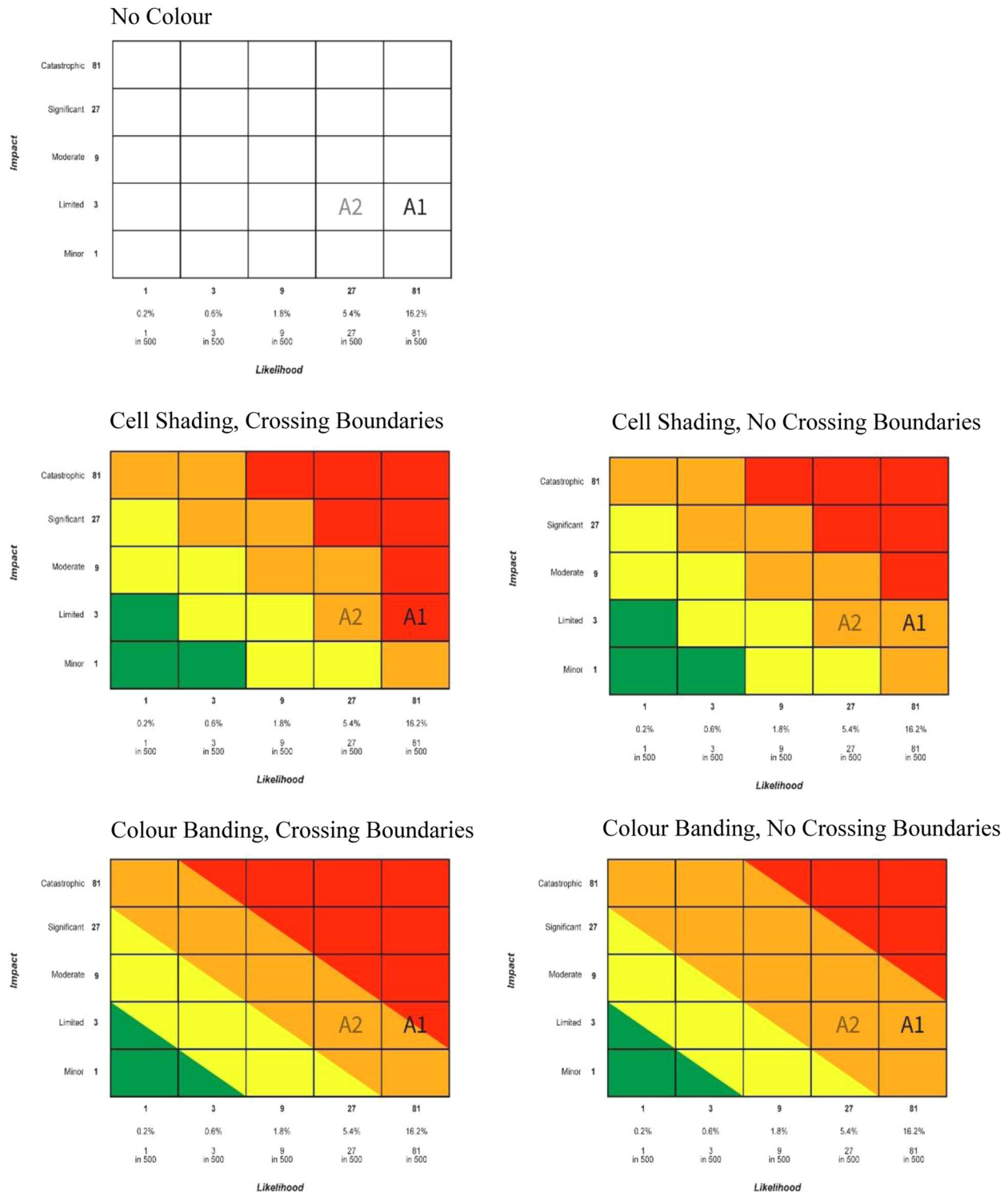


FIGURE 6 Example risk reduction stimuli used in the five arms of Experiment 1.2. Participants were asked how much they would be willing to pay to reduce risk A from position A1 to position A2.

($B = 18.357, t(4376.09) = 39.67, p < 0.001$) probability levels (large reductions). Similarly, participants were willing to pay less at medium probability levels than at high probability

levels ($B = 12.823, t(4372.85) = 27.80, p < 0.001$), indicating that the manipulation was successful. Unstandardized WTP betas indicate differences in thousands of pounds.

TABLE 2 Multilevel model structure—Predicting willingness-to-pay (WTP) at each probability level

	Unconditional model	Covariate model	Main effects model	Interaction model
Level 2: Participant	Random intercept	Random intercept	Random intercept	Random intercept
Numeracy	–	Fixed effect	Fixed effect	Fixed effect
Format	–	–	Fixed effect	Fixed effect
Format × Numeracy	–	–	–	Fixed effect
Level 2: Probability	Random intercept	Random intercept	Random intercept	Random intercept
Level 1: Item				
WTP amount (DV)				
ICC participant	0.486	0.481	0.479	0.477
ICC probability	0.141	0.143	0.143	0.143
Deviance change (X^2)	–	39.887***	12.809*	13.822**
AIC	58,264	58,227	58,214	58,204
BIC	58,291	58,261	58,275	58,292

Note: Tests for change in deviance were carried out using models fitted with ML, and models were compared from left to right, in the order shown. Fit indices are based on estimation with REML. AIC - Akaike Information Criterion; BIC - Bayesian Information Criterion; ICC - Intraclass Correlation Coefficient

* $p < 0.05$;

** $p < 0.01$;

*** $p < 0.001$.

3.2 | Results

3.2.1 | Analytical approach

Two plausible analytical approaches exist. One possibility is for the dependent variable to be the average of all WTP amounts across probability levels, predicted using a linear model. Another possibility is for the dependent variable to be represented by individual WTP amounts, and probability level to be included as a random effect in a multilevel model wherein individual WTP amounts are nested within individuals and within probability levels in a crossed random effects structure.

Originally, we pre-registered a mixed ANCOVA to study the effect of boundary crossing and color format on WTP amounts. However, we have used a multilevel model to enhance power. We have a sample size that is much lower than originally planned, and we have missing observations on our WTP measure, 67 at low probability levels, 45 at medium probability levels, and 36 at high probability levels. Multilevel models make use of all available data in estimation, whereas traditional linear models (e.g., mixed ANCOVA) would result in a loss of power as they require listwise deletion of cases. However, results from a nonhierarchical linear model predicting WTP amounts averaged across probability levels are presented in [Supporting Information Appendix D](#), and are also discussed below.

We started by partitioning overall variance in WTP amounts into participant-level, probability-level, and residual variance by including random intercepts for participants and probability levels. Subsequently, we added fixed effects for covariates, but only numeracy was found to minimize the deviance of the overall model meaningfully, and thus it was the only covariate considered. Table 2 illustrates the hierar-

chical structure of the data and shows the specification of the unconditional and final models, together with model fit indices, deviance statistics, and variance components. Including an interaction between format and numeracy achieved a reduction in deviance ($X^2(4) = 13.822$, $p = 0.008$) and AIC, so we decided to keep the interaction term despite the increase in BIC. Results from a model not including an interaction term are presented in [Supporting Information Appendix E](#) and are also discussed here. Additional exploratory analyses are presented in [Supporting Information Appendix F](#), but are not discussed here.

3.2.2 | Findings

Results of simple slopes analyses are presented in Table 3.

Cell shading groups

Crossing color boundaries versus no color. Participants in the boundary crossing condition were willing to pay more than participants in the no-color group, but only at high numeracy levels. However, this effect disappeared following alpha correction. A similar difference was found in the model predicting average WTP amounts at high numeracy levels prior to alpha adjustment, but not in the multilevel model without the interaction between format and numeracy. The effect therefore does not appear to be robust, failing to support H4a.

Cell shading groups: Crossing color boundaries versus no crossing color boundaries

Those in the boundary crossing condition were willing to pay more than participants for whom risk reductions did not cross color boundaries, with this effect increasing as numeracy increased, and reaching significance at average and high numeracy levels. This effect was also found by the model

TABLE 3 Predicting willingness-to-pay (WTP) at each probability level at low, mean, and high numeracy levels

Cell shading					
Reference level					
No color	Crossing boundaries				
Numeracy	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	–2.402	1.743	–1.378	0.168	0.253
Mean	0.666	1.251	0.532	0.595	0.714
High	3.734	1.785	2.092	0.037	0.073
No crossing boundaries	Crossing boundaries				
Numeracy	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	2.439	2.022	1.206	0.228	0.304
Mean	3.778	1.420	2.660	0.008	0.036
High	5.117	2.016	2.538	0.011	0.036
Color banding					
Reference level					
No color	Crossing boundaries				
Numeracy	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	–3.363	1.713	–1.963	0.050	0.085
Mean	0.479	1.220	0.393	0.695	0.758
High	4.321	1.716	2.519	0.012	0.036
No crossing boundaries	Crossing boundaries				
Numeracy	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>p</i> (B–H)
Low	–0.458	1.996	–0.229	0.819	0.819
Mean	2.991	1.420	2.106	0.035	0.073
High	6.441	2.016	3.196	0.001	0.017

Note: *p* (B–H)—value adjusted with the Benjamini–Hochberg procedure. *B*—unstandardized regression weights (unit = £1k in a range from £0 to £100k).

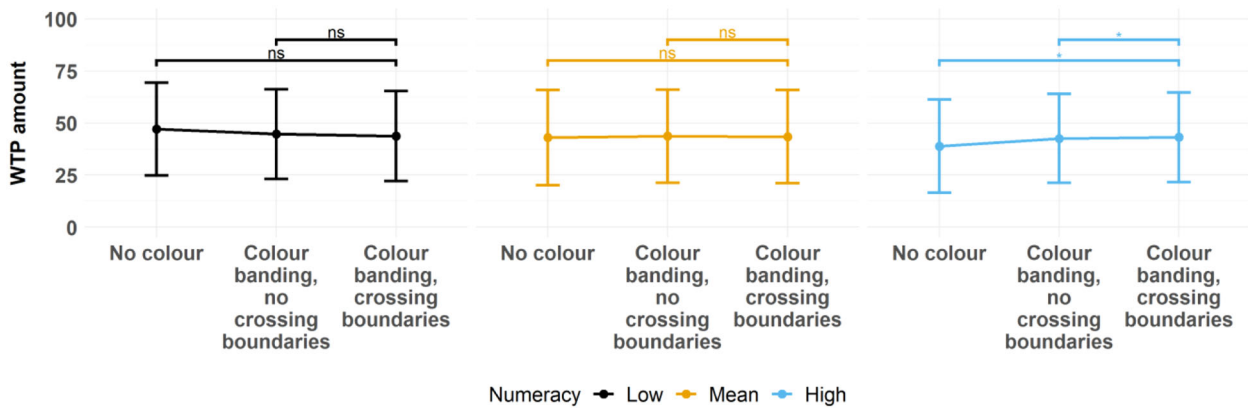


FIGURE 7 Marginal mean willingness-to-pay (WTP) amount and 95% confidence intervals in each format at different levels of numeracy. Significance levels: ns—*p* > 0.05, *—*p* < 0.05

predicting average WTP amounts at mean and high numeracy levels, and by the multilevel model without the interaction between numeracy and format. Therefore, H4b is supported, and the boundary crossing effect appears to increase with numeracy. The differences between formats at low, mean, and high numeracy levels are also illustrated in Figure 7.

Color banding groups: Crossing color boundaries versus no color

Those in the boundary crossing condition were willing to pay more than participants in the no-color group, but only at high numeracy levels. This effect was also found by the model predicting average WTP amounts at high numeracy levels, but

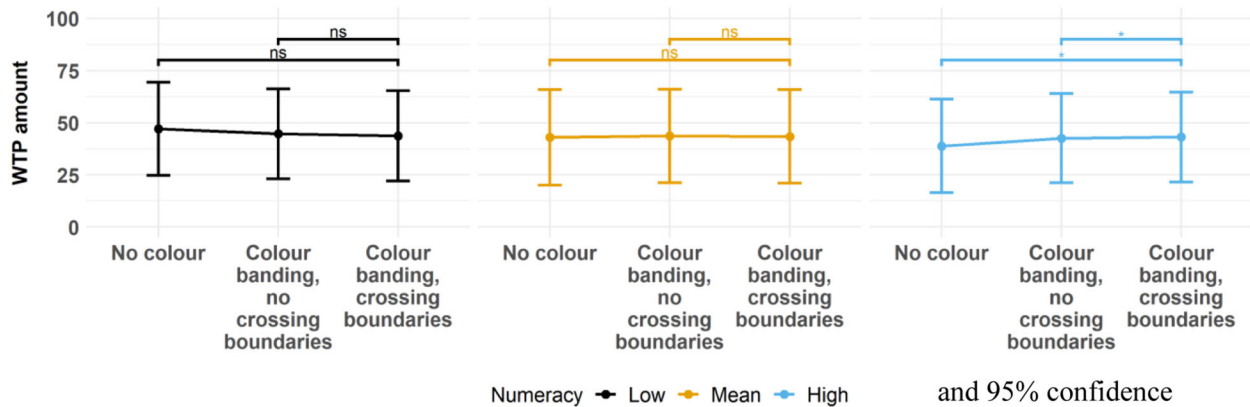


FIGURE 8 Marginal mean willingness-to-pay (WTP) amount and 95% confidence intervals in each format at different levels of numeracy, adjusted for flooding risk perceptions. Significance levels: ns— $p > 0.05$, *— $p < 0.05$

not by the multilevel model without the interaction between numeracy and format. Therefore, H4c is supported, with the effect occurring only at high numeracy levels.

3.2.2.4 Crossing color boundaries versus no crossing color boundaries

Those in the boundary crossing condition were willing to pay more than participants for whom risk reductions did not cross color boundaries, with this effect increasing as numeracy increased, and reaching significance at average and high numeracy levels. However, only the effect at high numeracy levels remained significant following alpha adjustment. The same pattern was found by the model predicting average WTP amounts, but the multilevel model without the interaction between numeracy and format only found this effect significant prior to alpha-level adjustment. Therefore, hypothesis H4d appears to be supported, and the boundary crossing effect holds only at high numeracy levels. The differences between formats at low, mean, and high numeracy levels are also illustrated in Figure 8.

4 | DISCUSSION

In these two experiments, we empirically investigated the potential for bias induced by color boundaries between cells in two risk matrix formats, with hypotheses consistent with fuzzy-trace theory (Reyna & Brainerd, 1995). We expected to find that colors in a risk matrix—conveying an ordinal description of risk severity—lead to reliance on ordinal, gist representations of risk information over precise numerical information, as proposed by the theory (Reyna & Brust-Renck, 2020). These decisions would sometimes favor risk reductions that cross color boundaries even if they achieve lower absolute risk reductions (i.e., the boundary-crossing effect). Our results only partially supported our hypotheses. Generally, we found evidence of a boundary-crossing effect more consistently in the color banding format (H2, H4c, H4d were supported) than in the cell shading format (H4b was supported, but H1, H3, and H4a were not). Perhaps most

surprisingly, we found that the effect appears to be stronger, or more often present, in individuals with higher numeracy. Results from the two studies are discussed below together with suggestions for future research.

4.1 | The boundary-crossing effect

We hypothesized that participants randomized to either the cell shading or the color banding format would be more likely to prefer a risk reduction which crossed color boundaries than participants allocated to the no-color condition, where the same risk reduction did not cross any boundary (Experiment 1: H1, H2). We also hypothesized that within each color format, participants for whom risk reductions crossed color boundaries would value the reductions more than participants for whom the same risk reductions did not cross color boundaries, or participants shown the same risk reductions in blank matrices (Experiment 2: H4a, H4b, H4c, H4d).

We found consistent support for these hypotheses in the color banding format, but not the cell shading format. Compared to participants in the no-color condition, participants in the color banding condition had lower risk comparison scores, that is, displayed the hypothesized boundary-crossing effect, but only at average and high numeracy levels (Experiment 1). Similarly, participants in the color banding condition for whom risk reductions crossed color boundaries were willing to pay more than participants for whom risk reductions did not cross color boundaries, and participants for whom risk reductions were presented in blank matrices, with these effects holding only at high numeracy levels (Experiment 2). This pattern gives credence to the reasoning that matrix coloration could influence ordinal gist perceptions, and bias participants into making and valuing decisions that might not be optimal—at least where the participants were of higher numeracy.

By contrast, there was no difference between the cell shading group and the no-color group in risk comparison scores, indicating that decision-making in the cell shading group was not influenced by the boundary-crossing

effect, contrary to our hypothesis (Experiment 1). Similarly, although participants in the cell shading format for whom risk reductions crossed color boundaries were willing to pay more than participants for whom risk reduction did not cross color boundaries at high levels of numeracy, differences from participants for whom the same risk reductions were presented in blank matrices were not significant (Experiment 2). Therefore, it is not clear whether the significant effect found in the second experiment is truly attributable to differences in the coloration, or is a false positive result.

It should be noted that where significant differences were found, the effects were very small, which could be attributed to the simplicity of our tasks. It might be that the boundary-crossing effect is more prominent in complex decision tasks (e.g., where participants have to compare more than two risks). Future research should employ more complex decision tasks, as these might more closely resemble the reality of decision-making using risk matrices, offering a more ecologically valid perspective on the boundary-crossing effect.

4.2 | Cell Shading versus color banding

We originally hypothesized that the increased granularity of risk severity categories displayed in the color banding format would lessen the boundary-crossing effect by reducing the perceived discrepancy between risk categories. Our results showed the reverse pattern, as participants in the color banding format were more susceptible to the boundary-crossing effect than participants in the cell shading format, but only at mean and high numeracy levels. It might be that increasing the granularity of risk categories had the opposite effect to the intended one and strengthened the ordinal gist representations induced by colors, with risk matrices having “half one color, half another” cells more strongly suggesting an ordinal transition between risk categories than a matrix wherein each cell is assigned one color only. It might be that the color banding format communicates ordinal transitions between risk categories more clearly than the cell shading format, and in so doing encourages participants to rely on ordinal gist representations in decision making. This explanation is consistent with the precept of fuzzy trace theory that the cognitive representation individuals rely on in decision-making depends on stimulus characteristics (Reyna & Brust-Renck, 2020).

4.3 | The moderating effect of numeracy

One particularly unexpected finding in our studies was that the boundary-crossing effect seemed to be present only, or increase for, individuals with average or high levels of numeracy. This pattern is consistent with other studies which find that highly numerate individuals display more biased decision-making than less numerate individuals (e.g., Kleber et al., 2013; Peters et al., 2019). A key precept of fuzzy-trace theory is that experts rely on gist knowledge when

making decisions, unlike novices, who rely more on verbatim information (Reyna et al., 2009). High numeracy can be interpreted as evidence of expertise with numbers, which thus should result in an increased reliance on gist when making decisions based on numeric information. Accordingly, in our study, individuals higher in numeracy were more likely to rely on ordinal gist processing when making and evaluating risk reductions, thus explaining their higher likelihood of displaying the boundary-crossing effect. This argument has been conceptually discussed in previous research (Peters et al., 2019), but it remains a speculative interpretation of the present findings since we did not include measures of ordinal gist in our study. Distinguishing between individual preferences for gist representations of numeric information (i.e., categorical and ordinal) and including these as predictors alongside objective numeracy could help test this argument. This approach has been used in previous research to explain decision paradoxes (e.g., the Allais paradox) and predict risk comparisons beyond the effect of objective numeracy (Reyna & Brust-Renck, 2020). Additionally, there is evidence that individuals low in numeracy might struggle to interpret risk information and therefore overestimate the likelihood of risks (Reyna et al., 2009). Indeed, in our study, there was some indication that participants higher in numeracy were willing to pay less for risk reductions, which might support the idea that participants lower in numeracy were willing to pay more because they overestimated the potential likelihood and impact of risks. It might be that for these participants the presence of color boundaries did not influence the amounts they were willing to pay due to a general difficulty in understanding the risk information presented to them.

4.4 | Limitations and future directions

The findings of this study should be interpreted in light of several limitations. First, exactly which cells appeared in which colors within the matrix sometimes had to change to make it possible for the same risk reduction to cross or not cross a color boundary, meaning the proportion of colors appearing in matrices was different across conditions. This was particularly problematic in the color banding format, and might have been at least partially responsible for our results. Second, these results are not generalizable to quantitative or qualitative risk matrices, as we only investigated the boundary-crossing effect in semiquantitative risk matrices. The complete absence of numeric information on the impact axis, or alternatively a fully quantitative impact axis, could lead participants to employ different cognitive processes beyond those discussed in this study. Third, our exploratory analyses (see [Supporting Information Appendix F](#)) failed to replicate effects in Experiment 2 when risk reductions were presented at low and medium probability levels, and when risk reductions were presented at impact level 1. This could point toward impact as a moderator of the boundary-crossing effect, such that this effect gets stronger as the

impact of the risk increases. To check the robustness of the boundary-crossing effect, future research should focus on systematically manipulating the impact levels at which risk reductions are presented. Finally, these findings were found in a sample of individuals wherein the majority had very little or no experience with risk matrices, but were asked to use them to make decisions about flooding risk mitigation. It might be that domain specialists (e.g., disaster risk reduction specialists) who are trained in the use, or contribute to the design of risk matrices do not display the same bias, so these findings might not generalize to them. Future research should investigate the extent to which individuals who have domain-specific expertise and training in the use of risk matrices are susceptible to the boundary-crossing effect. Future research might also benefit from investigating alternative ways of using color in risk matrices beyond the formats considered in this study, with a view to understanding whether and under what conditions they might lead to biases in decision making.

4.5 | Conclusion

These experiments show some preliminary evidence that the color assignment in risk matrices might influence people's perception of risk gravity, and therefore their decision-making with regards to risk mitigation. We found that individuals might be tempted to cross color boundaries when reducing risks even if this option is not advantageous (i.e., the boundary crossing effect). However, this effect was not consistently found when we included exploratory analyses of risk mitigations at different impact levels.

Pending future research replicating these results, the cautious recommendation is that the potential biasing effects of color should be considered alongside the goal of communication. If the purpose of communication is informing individuals in an unbiased way, these findings suggest it might be worth eliminating colors from risk matrices in order to reduce the risk of the boundary-crossing effect. On the other hand, if the goal of communication is to persuade individuals to implement certain risk mitigation actions, it might be that assigning colors so as to elicit the boundary-crossing effect would facilitate this. This could be the case, for example, when designing risk matrices that communicate action standards (i.e., severity level at which risk mitigation should be implemented) (Keller et al., 2009). This advice might be particularly relevant in the case of semiquantitative risk matrices, where color assignment might be arbitrary due to the absence of clear numeric cut-off points separating risk severity categories, and to situations where the users of the risk matrix are expected to be of higher numeracy and not have prior training in the design and use of risk matrices.

ACKNOWLEDGMENTS

This work was funded by the Winton Centre for Risk & Evidence Communication, which is supported by a donation from the David & Claudia Harding Foundation. We would also like to thank all the participants, those who

helped administrate the surveys, and to the anonymous peer reviewers who contributed to enhancing the quality of the manuscript.

ORCID

Ruri Proto  <https://orcid.org/0000-0002-9969-2136>

REFERENCES

- Cokely, E. T., Galesic, M., Ghazal, S., Garcia-Retamero, R., & Schulz, E. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7(1), 25–47. <http://journal.sjdm.org/11/11808/jdm11808.pdf>
- Cokely, E. T., Ghazal, S., Galesic, M., Garcia-Retamero, R., & Schulz, E. (2013). How to measure risk comprehension in educated samples. In R. Garcia-Retamero & M. Galesic (Eds.), *Transparent communication of health risks: Overcoming cultural differences* (pp. 29–52). Springer Science+Business Media. <https://doi.org/10.1007/978-1-4614-4358-2>
- Cotti, D., Harb, M., Hadri, A., Aboufirass, M., Chaham, K. R., Libertino, A., Campo, L., Trasforini, E., Krätzschar, E., Bellert, F., & Hagenlocher, M. (2022). An integrated multi-risk assessment for floods and drought in the Marrakech-Safi region (Morocco). *Frontiers in Water*, 4, (June), 1–17. <https://doi.org/10.3389/frwa.2022.886648>
- Cox, L. A. (2008). What's wrong with risk matrices? *Risk Analysis*, 28(2), 497–512. <https://doi.org/10.1111/j.1539-6924.2008.01030.x>
- Cox, L. A., Babayev, D., & Huber, W. (2005). Some limitations of qualitative risk rating systems. *Risk Analysis*, 25(3), 651–662. <https://doi.org/10.1111/j.1539-6924.2005.00615.x>
- Duijm, N. J. (2015). Recommendations on the use and design of risk matrices. *Safety Science*, 76, 21–31. <https://doi.org/10.1016/j.ssci.2015.02.014>
- He, Z., & Zhai, G. (2017). How does probability judgment influence contingent valuation method to estimate WTP for natural disaster reduction? *Human and Ecological Risk Assessment*, 23(3), 508–518. <https://doi.org/10.1080/10807039.2016.1208048>
- Huet, E. K., Adam, M., Traore, B., Giller, K. E., & Descheemaeker, K. (2022). Coping with cereal production risks due to the vagaries of weather, labour shortages and input markets through management in southern Mali. *European Journal of Agronomy*, 140, (July), 126587. <https://doi.org/10.1016/j.eja.2022.126587>
- International Organization for Standardization. (2019). Risk Management – Risk Assessment Techniques (ISO Standard No. 31010: 2019). <https://www.iso.org/standard/72140.html>
- Kadkhodaei, M. H., & Ghasemi, E. (2022). Development of a semi-quantitative framework to assess rockburst risk using risk matrix and logistic model tree. *Geotechnical and Geological Engineering*, 40(7), 3669–3685. <https://doi.org/10.1007/s10706-022-02122-9>
- Keller, C., Siegrist, M., & Visschers, V. (2009). Effect of risk ladder format on risk perception in high- and low-numerate individuals. *Risk Analysis*, 29(9), 1255–1264. <https://doi.org/10.1111/j.1539-6924.2009.01261.x>
- Kleber, J., Dickert, S., Peters, E., & Florack, A. (2013). Same numbers, different meanings: How numeracy influences the importance of numbers for pro-social behavior. *Journal of Experimental Social Psychology*, 49(4), 699–705. <https://doi.org/10.1016/j.jesp.2013.02.009>
- Lemmens, S. M. P., Lopes van Balen, V. A., Röselaers, Y. C. M., Scheepers, H. C. J., & Spaanderman, M. E. A. (2022). The risk matrix approach: A helpful tool weighing probability and impact when deciding on preventive and diagnostic interventions. *BMC Health Services Research*, 22(1), 1–11. <https://doi.org/10.1186/s12913-022-07484-7>
- Levine, E. S. (2012). Improving risk matrices: The advantages of logarithmically scaled axes. *Journal of Risk Research*, 15(2), 209–222. <https://doi.org/10.1080/13669877.2011.634514>
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1), 37–44. <https://doi.org/10.1177/0272989x0102100105>
- Monat, J. P., & Doremus, S. (2020). An improved alternative to heat map risk matrices for project risk prioritization. *Journal of Modern Project Management*, 7(4), 214–228. <https://doi.org/10.19255/JMPM02210>

- Mu, D., Kaplan, T. R., & Dankers, R. (2018). Decision making with risk-based weather warnings. *International Journal of Disaster Risk Reduction*, 30, (March), 59–73. <https://doi.org/10.1016/j.ijdr.2018.03.030>
- Peters, E., Fennema, M. G., & Tiede, K. E. (2019). The loss-bet paradox: Actuaries, accountants, and other numerate people rate numerically inferior gambles as superior. *Journal of Behavioral Decision Making*, 32(1), 15–29. <https://doi.org/10.1002/bdm.2085>
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7(1), 1–75. [https://doi.org/10.1016/1041-6080\(95\)90031-4](https://doi.org/10.1016/1041-6080(95)90031-4)
- Reyna, V. F., & Brust-Renck, P. G. (2020). How representations of number and numeracy predict decision paradoxes: A fuzzy-trace theory approach. *Journal of Behavioral Decision Making*, 33(5), 606–628. <https://doi.org/10.1002/bdm.2179>
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135(6), 943. <https://doi.org/10.1037/a0017327>
- Roman, L., Hardesty, B. D., & Schuyler, Q. (2022). A systematic review and risk matrix of plastic litter impacts on aquatic wildlife: A case study of the Mekong and Ganges River Basins. *Science of the Total Environment*, 843, (June), 156858. <https://doi.org/10.1016/j.scitotenv.2022.156858>
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127(11), 966–972. <https://doi.org/10.7326/0003-4819-127-11-199712010-00003>
- Sutherland, H., Recchia, G., Dryhurst, S., & Freeman, A. L. J. (2021). How people understand risk matrices, and how matrix design can improve their use: Findings from randomized controlled studies. *Risk Analysis*, 42(0), 1023–1041. <https://doi.org/10.1111/risa.13822>
- The UK Cabinet Office. (2017). National Risk Register of Civil Emergencies. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211867/NationalRiskRegister2013_amended.pdf
- World Economic Forum. (2021). *The global risks report 2021* (16th Ed.). World Economic Forum.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Proto, R., Recchia, G., Dryhurst, S., & Freeman, A. L. J. (2022). Do colored cells in risk matrices affect decision-making and risk perception? Insights from randomized controlled studies. *Risk Analysis*, 1–15. <https://doi.org/10.1111/risa.14091>