Misperception of Multiple Risks in Medical Decision-Making

MEHDI MOURALI
ZHIYONG YANG

How do consumers combine multiple risk items when forming overall risk judgments? Drawing on the fuzzy trace theory and categorical averaging, this research finds that adding a low-risk item to a high-risk item reduces the overall risk perception because people reason categorically about risk. They impose categorical distinctions on quantitative risk information, and when combining categorical information, they tend to average across categories instead of adding. Eight studies in the context of medical and health product decisions ($N = 5,152$) provide convergent evidence showing that when items in different risk tiers are considered together, they are consistently evaluated in a subtractive manner, leading to a higher likelihood of taking the objectively riskier medication (studies 1A, 1B, 1C, and 2). This effect is eliminated when the probability of one side effect is exceedingly high (study 3) or when the task requires reliance on verbatim representation of probabilities (study 4). The effect also disappears when risk information is presented graphically in a way that emphasizes the additive property of multiple risks (studies 5A and 5B). The findings have important implications for the fields of risk perception, risk communication, and consumer health and medical decision-making.

Keywords: risk perception, risk communication, categorical thinking, fuzzy trace theory, medical decision-making, graphical representation

Last year, Joanne was diagnosed with rheumatoid arthritis. Initially, she responded well to her treatment but, in recent months, her arthritis has flared up again. Her doctor tells her about a new medication that may help her. The doctor explains that the new medication is effective for about 75% of people who take it. However, it comes with some risk of side effects. Fifty-five percent of people who take it experience abdominal cramps and pain, and 6% experience blurry vision and increased light sensitivity. After consulting with her doctor, Joanne needs to make the final call on whether to start the new medication or continue with her existing treatment. Just like Joanne, consumers are increasingly expected to play an active role in their health and medical care decisions (Aydin and Gokcen 2019), including collaborating with physicians to define and implement care plans (Hibbard 2004). However, for consumers to make informed decisions, they must clearly understand the risks and benefits associated with different actions.

Health and medical risk information is commonly encountered in the form of numbers, though verbal and graphical formats are also used occasionally (Lipkus 2007). For example, consumers may learn that they have a 10% chance of contracting a disease or that 1 in 100 people taking a certain medication will experience an adverse effect. Understanding how consumers perceive and process numerical risk information is not only of theoretical interest, but it also has important implications for designing effective risk communications and facilitating informed decision-making.

Mehdi Mourali (mehdi.mourali@haskayne.ucalgary.ca) is an associate professor of marketing at the Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada. Zhiyong Yang is a professor of marketing at the Farmer School of Business, Miami University, 800 E. High St. Oxford, OH 45056, USA. Please address correspondence to Mehdi Mourali. This research was supported by a Haskayne Research Professorship awarded to the first author. Supplementary materials are included in the web appendix accompanying the online version of the article.

Editors: Margaret C. Campbell and Stacy Wood

Associate Editor: Nailya Ordabayeva

Advance Access publication September 1, 2022

 • Vol. 50 • 2023
https://doi.org/10.1093/jcr/ucac040

A vast literature going back several decades has investigated risk perception, including its antecedents, moderators, and consequences (Slovic 2000). In the medical decision-making domain, the lion's share of this work has focused on the perception of single risks. Less is known about how consumers combine multiple risk items when forming overall risk impressions (Leonhardt and Keller 2018). For example, consider a medication with multiple potential adverse effects. How do consumers combine information about the probabilities of these adverse effects when assessing the medication's overall risk?

Normatively, judgments of multiple, mutually exclusive risks ought to be additive. That is, consumers should judge an option with multiple potential adverse effects to be riskier than an option with only one of these potential adverse effects. A medication that comes with one side effect occurring in 30% of the people who take it and a second side effect (SSE) affecting 1% of the people who take it is objectively riskier than a medication that comes with only the first side effect (FSE). However, in sharp contrast to this fact, we show that consumers judge the medication with one high-probability and one low-probability side effect to be less risky than the medication with only one high-probability side effect. We posit that adding a low-probability side effect reduces the overall risk perception because people reason categorically about risk. They tend to impose categorical distinctions on quantitative risk information, and when combining categorical information, they tend to average across categories instead of adding.

The resulting distortion in the perception of multiple risks from different risk categories can have serious downstream consequences for consumers' welfare. For example, we find that adding a low-risk side effect to a medication associated with a high-risk side effect increases consumers' likelihood to start the medication. That is, consumers are more willing to take the objectively riskier drug because they misperceive it to be safer. We show that these effects disappear when judgments are based on the verbatim probability of side effects or when the risk of one side effect is perceived to be too high. We also show that presenting risk information graphically in a way that emphasizes the additive property of multiple risks reduces both the distortion in risk perception and its impact on consumers' willingness to start the medication.

THEORETICAL BACKGROUND

Consumers encounter multiple risk options in a variety of health contexts, such as when evaluating medical treatments, surgery, screening, or immunization. For example, the World Health Organization[1] reports that the live attenuated influenza vaccine has the following possible side effects: runny nose/nasal congestion (59–63%), cough (28%), fever (16–31%), decreased activity (16–23%), vomiting (10%), abdominal pain (4%), muscle aches (14%), wheeze in children of 6–11 months of age (14%), and anaphylaxis (1 per 500,000).

Extant research shows that the severity of the various side effects impacts consumers' risk perceptions. For instance, adding less severe side effects (e.g., congestion and fatigue) to the description of a drug that is associated with a more severe side effect (seizures) can reduce the perceived negativity of the drug (Khan and Kupor 2017). Presumably, this happens because the presence of minor prospects makes the major prospect seem less likely. Listing minor side effects along with severe side effects was also found to dilute the overall perceived severity of a drug's side effects and to increase its attractiveness relative to a drug that only lists the severe side effects (Sivanathan and Kakkar 2017). These studies are highly instructive about the impact of side effect severity on risk perception. However, risk judgments depend not just on the severity but also on the probabilities of the various side effects. In particular, the way consumers integrate information about the probabilities of multiple side effects can influence their risk perceptions over and above any impact of perceived severity.

Other research investigating the perception of multiple risks has generally reported subadditive effects (French et al. 2000; Hermand, Mullet, and Coutelle 1995). For instance, both smoking and a family history of heart attacks are considered high-risk factors for heart attacks. Yet, a man described as having both risk factors was judged to be only slightly more likely to have a heart attack than a man described as having only one of the risk factors (French et al. 2000). In contrast, we argue that the presence of multiple risks can have subtractive (instead of subadditive) effects on overall risk perceptions. Our prediction is based on insights from fuzzy trace theory and categorical reasoning.

Fuzzy Trace Theory

Fuzzy trace theory (Reyna and Brainerd 1991, 1995) distinguishes between verbatim and gist mental representations of information. Verbatim representation is precise and captures the exact surface form of the information. In contrast, gist representations are less precise and encapsulate the bottom-line meaning, or subjective interpretation of the information. When consumers encounter quantitative information, they encode both the verbatim representation of the numbers and multiple gist representations that capture their interpretations of the numbers. For example, when told that 1% of people taking medication A experience headaches and 30% of people taking medication B experience headaches, consumers encode the verbatim risk

1 https://web.archive.org/web/20211210101221/https://www.who.int/vaccine_safety/initiative/tools/Influenza_Vaccine_rates_information_sheet.pdf?ua=1

levels of 1% and 30%, and, in parallel, they also encode such gists as "headaches are rare with medication A," "headaches are quite common with medication B," "headaches are more likely with medication B than with medication A," and so on.

The theory posits that people prefer to rely on the least precise representation that still allows them to successfully complete their task. That is, they start with the most basic representation and then proceed to more precise representations as the task demands. For example, when assessing an option's riskiness, a basic gist is the categorical distinction between "no risk" and "some risk," or between "low risk" and "high risk." However, such representations are insufficient when the task requires estimating exact probabilities. In this case, the more precise verbatim representation (e.g., "headaches happen for 30% of the people taking medication B") is used instead.

Fuzzy trace theory provides a parsimonious explanation for many classical judgment and decision-making phenomena, such as framing effects (Corbin et al. 2015; Reyna and Brainerd 1991), overestimation of small risks (Reyna 2004), ratio bias (Reyna and Brainerd 1994), and base rate neglect (Reyna and Adam 2003). Indeed, many of these biases seem to originate in the difficulties people have in translating numerical information into meaningful representations. Crucially, support for the parallel encoding of gist and verbatim representations and for people's general inclination to rely on the crudest gist representation that accomplishes the job is wide-ranging and includes evidence from controlled experiments, mathematical modeling, and neuroimaging studies (see Reyna 2008 for a review). Fuzzy trace theory, thus, provides a strong theoretical foundation for our premise that consumers tend to classify quantitative risk information into discrete categories such as "low risk" and "high risk" and prefer to operate on these categorical judgments when integrating information about multiple risks. Further support for this premise is found in broader research on categorization showing that people often think in categories to simplify complex information and facilitate decision-making (Gutman 1982; Murphy and Ross 1994; Rozin, Ashmore, and Markwith 1996; Smith and Medin 1981), and in a pilot study, we designed to explore whether consumers spontaneously categorize quantitative risk information. Participants in the pilot study ($N = 216$) read that 45% of people who take a new arthritis medication experience abdominal cramps and pain. They were then asked to complete the following sentence: "The risk of experiencing side effects when taking this medication is . . ." Their answers were coded by two independent judges (inter-rater agreement = 94.4%) as reflecting either a numerical judgment of risk (e.g., "45%," "forty-five percent"), a categorical judgment of risk (e.g., "high," "fairly likely"), or neither (e.g., "abdominal cramps," "worth it"). Most responses (50.5%) reflected categorical risk judgments, while 25.9% reflected numerical risk judgments, and 23.6% reflected neither. These results are consistent with the fuzzy trace theory and the notion that consumers tend to spontaneously categorize quantitative risk information.

This line of theorizing raises an important question regarding the relation between numbers and categories: how exactly are numerical probabilities mapped onto risk categories? In the same pilot study, we asked participants to classify each of 10 different levels of risk probabilities ranging from 1% to 75% (i.e., 1%, 2%, 6%, 17%, 22%, 30%, 35%, 45%, 65%, and 75%) as either very low, low, moderate, high, or very high. The results (detailed in web appendix 4) indicate that risk probabilities in the 1–6% range were typically considered very low to low. Probabilities of 17% and 22% were typically classified as low to moderate risk. Probabilities in the 30–45% range were considered moderate to high risk, and probabilities of 65% and 75% were considered high to very high risk. These results are instructive, but it is important to note that the exact mapping of numerical probabilities onto risk categories depends on both the individual making the judgment and the context in which the risk is being judged. For example, Reyna and Brainerd (2008) noted that the gist of a quantity often depends on its relation to other quantities in the context. Participants in one study rated a 12% risk of disease in women as "high" when it was compared to a 4% risk in men. However, they rated the same 12% risk as "low" when it was compared to a 20% risk in men (Windschitl, Martin, and Flugstad 2002). Likewise, people exhibited a greater tendency to classify a 13% lifetime risk of breast cancer as "low" after providing comparatively higher initial estimates than in the absence of initial estimates (Fagerlin, Zikmund-Fisher, and Ubel 2005).

Having established that consumers mentally represent quantitative risk information in discrete categories and prefer to operate on the latter, we now turn to the question of how they combine multiple categorical judgments when assessing an option's overall riskiness.

Forming Overall Risk Judgments

Information integration theory (Anderson 1981) posits that people form overall impressions of stimuli by engaging in some sort of "cognitive algebra" whereby they use simple computational rules to integrate information from multiple cues. These computational rules are functionally equivalent to general-purpose adding, subtracting, multiplying, and averaging procedures. For example, when deciding among competing options, consumers may use an adding procedure, in which they simply tally unweighted cues and choose the option with more positive features (Alba and Marmorstein 1987; Dawes 1971). Alternatively, consumers may use an averaging procedure, in which attribute information is combined by averaging across attributes instead of adding. Averaging has the distinctive

property of producing subtractive judgments. That is, when consumers average across attributes, the resulting summary evaluation becomes less extreme than the evaluation of either the most positive or the most negative attribute. Prior research has shown that providing moderately positive information in conjunction with highly positive information results in less favorable overall evaluation, whereas providing moderately negative information in conjunction with highly negative information results in less unfavorable evaluations (Anderson 1965; Anderson and Alexander 1971; Shanteau 1975; Troutman and Shanteau 1976). It is worth noting that consumers may also avoid integrating information altogether by relying on a single attribute to quickly form a conclusion or make a decision. The recognition heuristic (Gigerenzer and Goldstein 1996, 2011) and the take-the-best heuristic (Gigerenzer, Todd, and the ABC Research Group 1999) are examples of cognitive strategies that rely on a single cue.

We propose that when combining items with categorically disparate risk levels, consumers tend to use an averaging rule. We expect that averaging across categorical judgments of "high risk" and "low risk" will lead to lower overall risk assessment for the combined items than for the "high risk" item alone. Our predictions build on prior work on categorial reasoning, showing a preponderance for averaging when integrating categorical judgments to form overall impressions. Such categorical averaging was found to underly consumer judgments in various settings, including perception of quality in multi-attribute products (Troutman and Shanteau 1976), assessment of the monetary value of product bundles (Brough and Chernev 2012), estimation of calories in meals (Chernev and Gal 2010), and evaluation of how likeable a person is based on their personality traits (Anderson and Alexander 1971). For instance, Brough and Chernev (2012) found that a bundle that includes one inexpensive item and one expensive item is valued less than the expensive item alone. Similarly, Chernev and Gal (2010) found that consumers typically underestimate the calories contained in a meal combining healthy and indulgent options because they are inclined to average rather than add the calories of the healthy and indulgent options.

The present investigation extends this line of research by showing that categorical averaging is also at the root of biased risk perception in medical decision-making. Importantly, it offers new insights on the averaging process and its boundary conditions. It demonstrates that consumers do not in fact average the numerical probabilities they encounter but the cruder risk categories that these probabilities represent. It also shows that averaging is less likely when the judgment task requires the use of verbatim representations of the probability information, or when the risk of one side effect is perceived to be too high. Finally, it proposes a scalable intervention for improving risk communication. It shows that presenting risk information graphically in a way that emphasizes the additive property of multiple risks mitigates both the distortion in risk perception and its impact on consumers' willingness to start the medication.

Boundary Conditions

We have argued that subtractive risk judgments happen because consumers prefer to rely on their categorical representations of the risk information instead of the actual probabilities and because they tend to average across categorical judgments when forming an overall impression. Thus, disrupting either consumers' reliance on categorical representations of risk or their tendency to average across categories should eliminate subtractive risk judgments.

Disrupting Reliance on Categorical Representation of Risk. Fuzzy trace theory suggests that the default mode of operation is to think about risk in terms of vague categories (i.e., the gist) rather than precise numbers (i.e., verbatim). However, since numbers are also encoded, people will use them when the task at hand does not permit the use of more basic gist (Reyna and Brainerd 2008). One such case is when consumers are explicitly asked to estimate probabilities (Reyna and Brainerd 2008). Thus, we predict that asking consumers to estimate the numerical probability of experiencing a side effect, as opposed to asking about overall risk, would reduce subtractive judgments.

Disrupting Categorical Averaging: When the Risk Is Too High. Even when consumers think categorically about risk, subtractive risk judgments would not occur if the averaging rule is not used to integrate information about multiple risks. A side effect that is perceived to be too risky (e.g., very high probability of occurring) is likely to have a disproportionate impact on the judgment of overall risk. In such cases, consumers are less likely to engage in categorical averaging. The presence of an exceedingly risky side effect offers a salient and compelling cue for judging an option's overall riskiness. As a result, consumers may not bother integrating risk information about the SSE. This is consistent with prior research showing that people often rely on a single salient cue when forming judgments and making decisions (Gigerenzer and Goldstein 1996; Newell and Shanks 2003). We expect the presence of a very risky side effect to act as such a single salient cue. Accordingly, we predict that subtractive risk judgments are less likely to occur when one side effect is perceived to be too risky.

Disrupting Categorical Averaging: Visual Aids. Categorical averaging can be disrupted not only when consumers rely on a single cue but also when they are encouraged to use an additive integration rule instead. We suggest that this can be achieved by means of suitable visualizations.

Visual aids are often recommended for enhancing risk communication (Ancker et al. 2006; Lipkus and Hollands

1999). Graphs have been shown to improve comprehension of risks associated with different medical treatments and screening tests (Okan et al. 2012; Waters et al. 2007; Zikmund-Fisher et al. 2008). They have also been found to amplify risk avoidance (Schirillo and Stone 2005) and increase consumers' willingness to pay for safer and healthier products (Stone et al. 2003; Stone, Yates, and Parker 1997). Graphical representations facilitate the processing of risk information. They attract and hold people's attention and prime automatic mathematical operations. For example, comparing the heights of two bars in a histogram automatically evokes subtraction rules (Lipkus 2007; Lipkus and Hollands 1999). However, not all graphs are equally effective at communicating risk information. Different graph types are better suited for different tasks. For instance, line graphs can effectively identify trends over time and enhance understanding of cumulative risk, whereas bar graphs and pictographs or icon arrays are more useful for making comparisons (Visschers et al. 2009). We propose that graphical representations designed to visually highlight the additive property of multiple risks will evoke an additive integration rule and reduce the distortion in the perception of multiple risks. Importantly, this effect is only expected for graphs that explicitly underscore the additive nature of multiple risks.

STUDY 1A: SINGLE RISK VERSUS COMBINED RISKS

Method

We examine our conjectures in the context of a decision to start a new medication that is associated with either one high-risk side effect or a combination of that high-risk side effect and another low-risk side effect. The studies were approved by the University of Calgary Conjoint Faculty Research Board and were conducted according to the principles expressed in the Declaration of Helsinki. In studies 1–4, we aimed to collect a minimum of 100 valid observations per cell. In study 5, we aimed for a minimum of 150 observations per cell, as we expected the graphical interventions to have a smaller effect. In all cases, we oversampled to account for potential loss of data due to failure to pass an attention check (web appendix 2). Our oversampling was greater in studies 1B and 1C to account for the additional screening in those studies (participants needed to have considered buying supplements/health and wellness products in the past year). The initial and final sample sizes (before and after exclusion) for all studies are summarized in web appendix 3.

The first experiment tests the prediction that consumers will perceive lower overall risk in the combined risks condition than in the single high-risk side effect condition. To enhance the validity of our findings, we also include a third condition consisting of a combination of two high-risk side effects. We expect that the two high-risk combinations will be perceived to be at a higher risk than either one high-risk side effect or a combination of that high-risk side effect and another low-risk side effect.

Three hundred and thirty US residents recruited from Prolific Academic took part in this experiment in exchange for monetary compensation. Participants were randomly assigned to one of three conditions (single vs. high–low vs. high–high). In the single condition, the new medication has a single side effect with a relatively high probability of occurrence. In the high–low condition, the medication is associated with a combination of the same side effect as in the single condition and an SSE with a low probability of occurrence. In the high–high condition, the medication is associated with a combination of the same side effect as in the single condition and an SSE with a high probability of occurrence. After excluding participants who did not pass the attention check, we were left with 295 participants (149 men, 142 women, 1 preferred not to say, and 3 identified as other; $M_{age} = 33.5$, $SD_{age} = 12.8$).

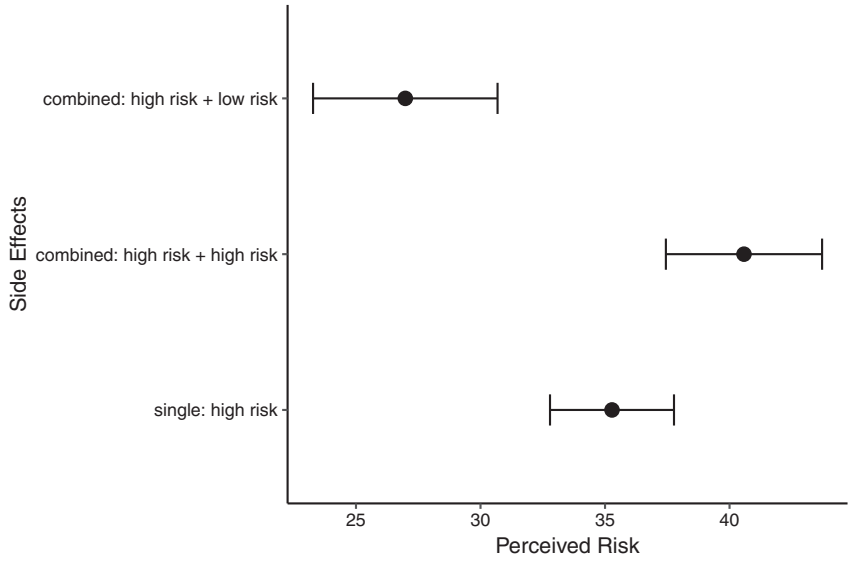
Participants were presented with a medical scenario adapted from Wilhelms, Fraenkel, and Reyna (2018), in which they imagined being diagnosed with rheumatoid arthritis (see web appendix 1 for a full description). Those in the single side effect condition learned that a new medication that may help them comes with a 30% chance of experiencing abdominal cramps and pain. Those in the combined high–low condition learned that in addition to the 30% chance of abdominal cramps and pain, there was a 1% chance of experiencing blurry vision and increased light sensitivity. Finally, those in the combined high–high condition read that in addition to the 30% chance of abdominal cramps and pain, there was a 35% chance of experiencing blurry vision and increased light sensitivity.

After reading the scenario, participants rated the overall risk of experiencing side effects when taking the medication on a 100-point sliding scale (0 = *very low risk*, 100 = *very high risk*), and their likelihood to start the medication, on a 100-point sliding scale (0 = *not likely at all*, 100 = *very likely*).

Next, participants completed the Berlin numeracy test (Cokely et al. 2012). Prior research (Keller and Siegrist 2009; Peters 2020; Reyna et al. 2009; Schwartz et al. 1997) suggests that low numeracy hinders people's understanding of risk information. Thus, we sought to control for the effect of numeracy in all our studies. The Berlin numeracy test (web appendix 2) consists of four multiple choice items designed to assess statistical numeracy and risk literacy. The following is a sample item: *Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent (a = 10%, b = 25%, c = 40%, d = None of the above).*

FIGURE 1

STUDY 1A: THE IMPACT OF SIDE EFFECTS ON PERCEIVED RISK



NOTE.—Bars in graph represent 95% confidence intervals.

Answers are coded 1 for a correct answer and 0 for an incorrect answer. The final numeracy score is the sum of all correct answers. Following the numeracy test, participants completed demographic measures and rated their experience in making decisions about medication on a 7-point scale (1 = *no experience at all*, 7 = *a lot of experience*).

Results and Discussion

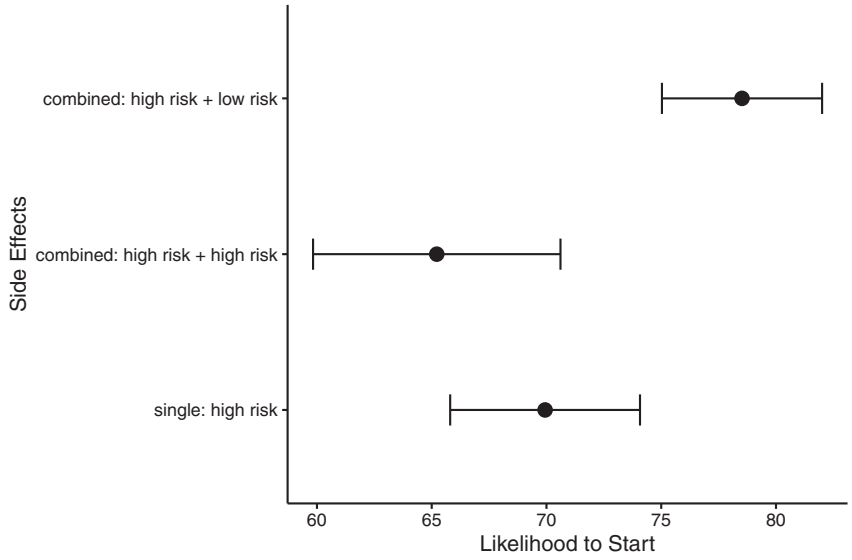
In all the studies, we analyzed the data both with and without covariates (numeracy and experience). In all cases, adding the covariates did not materially change the effects of the independent variables. For clarity of exposition and consistency of reporting, we only report analyses with the covariates in the main article. Analyses without the covariates are reported in web appendix 13. Data analysis was conducted using the statistical program R, version 4.1.1 (R Core Team 2021).

Perceived Risk. A one-way ANCOVA on perceived risk, using numeracy and experience with making medication-related decisions as control variables, found a significant effect of side effect condition ($F(2, 290) = 17.53$, $p < .001$, $\eta_p^2 = 0.11$). Planned contrasts with Bonferroni correction for multiple tests revealed that consumers who read about the combination of one high-risk side effect and one low-risk side effect ($M_{high–low} = 27.0$, $SD_{high–low} = 18.8$) reported lower overall risk than those who read about a single high-risk side effect ($M_{single} = 35.3$, $SD_{single} = 12.4$; $t(290) = -3.52$, $p < .001$). In contrast, consumers who read about the combination of two high-risk side effects perceived greater overall risk than those who read about the single high-risk side effect ($M_{high–high} = 40.6$, $SD_{high–high} = 16.1$; $t(290) = 2.29$, $p = .05$). These results are consistent with our theorizing and the prediction that averaging only occurs when the risks are perceived to be at qualitatively disparate levels. Furthermore, neither numeracy ($b = -0.56$, SE = 0.88; $F(1, 290) = 0.41$, $p = .52$, $\eta_p^2 = 0.001$) nor experience ($b = -0.59$, SE = 0.51; $F(1, 290) = 1.32$, $p = .25$, $\eta_p^2 = 0.004$) had a significant effect on perceived risk. Figure 1 illustrates the influence of side effects on perceived risk.

Likelihood to Start. A one-way ANCOVA found a significant effect of side effect condition ($F(2, 290) = 8.39$, $p < .001$, $\eta_p^2 = 0.055$). Participants in the combined high–low condition indicated a higher likelihood to start the medication ($M_{high–low} = 78.5$, $SD_{high–low} = 17.7$) than those in the single side effect condition ($M_{single} = 69.9$, $SD_{single} = 20.6$; $t(290) = 2.57$, $p = .02$). Those in the high–high condition indicated a lower likelihood to start the medication than those in the single side effect condition, though the difference here was not statistically significant ($M_{high–high} = 65.2$, $SD_{high–high} = 27.6$; $t(290) = -1.43$, $p = .31$). Moreover, neither experience with

FIGURE 2

STUDY 1A: THE IMPACT OF SIDE EFFECTS ON LIKELIHOOD TO START



NOTE.—Bars in graph represent 95% confidence intervals.

making medication decisions ($b = 1.22$, SE = 0.72; $F(1, 290) = 2.88$, $p = .09$, $\eta_p^2 = 0.010$) nor numeracy ($b = 0.32$, SE = 1.23; $F(1, 290) = 0.07$, $p = .79$, $\eta_p^2 = 0.000$) had a significant effect on likelihood to start the medication. Figure 2 illustrates the impact of side effects condition on likelihood to start the medication.

Mediation Analysis. We conducted a mediation analysis, focusing on the high–low and single side effect conditions. We examined the bias-corrected confidence intervals using 5,000 bootstrap iterations. The direct effect of scenario on likelihood to start the medication was not significant (ADE = 3.23, 95% CI [−1.62, 8.02]), but the indirect effect through perceived risk was significant (ACME = 5.03, 95% CI [2.20, 8.23]). The direction of the indirect effect confirmed that consumers who saw the combined side effects perceived lower overall risk than those who saw only one side effect, which, in turn, contributed to increasing their likelihood to start the medication.

Ruling Out Severity-Based Dilution as an Alternative Explanation. One might argue that subtractive judgments in the assessment of overall risk are driven not so much by categorical averaging of high and low probabilities of side effects but by a severity-based dilution of perceived risk (Sivanathan and Kakkar 2017). According to this account, adding an SSE that is perceived to be less severe than the FSE could lower overall risk perception and account for the observed subtractive judgments. If this was the case, however, we would expect the SSE to reduce the overall risk perception regardless of its probability of occurrence. Our data clearly do not support this prediction, as subtractive judgments only happened when the SSE had a low probability of occurrence relative to the FSE. Furthermore, data from our pretest show that the SSE (blurry vision and light sensitivity) was in fact seen as slightly more severe, not less severe, than the FSE (abdominal cramps and pain). Participants in the pretest ($N = 110$) rated the severity of each side effect on three 11-point items (1 = *not a big deal*, 11 = *a huge deal*; 1 = *not bad at all*, 11 = *very bad*; 1 = *mild*, 11 = *severe*). We constructed a severity index for abdominal cramps ($\alpha = 0.92$) and a severity index for blurry vision ($\alpha = 0.95$) by averaging scores across the three items for each side effect. A paired samples *t*-test found that blurry vision was considered more severe ($M = 8.00$, SD = 2.39) than abdominal cramps ($M = 7.51$, SD = 2.12; $t(109) = 2.17$, $p = .03$). Therefore, a severity-based dilution of perceived risk is unlikely to account for our findings.

STUDY 1B: REPRLICATION USING AN INCENTIVE-COMPATIBLE TASK

Study 1B tests the robustness of the previous findings using a different product, a different population, and an

incentive-compatible decision task. Moreover, by manipulating which side effect has a high probability of occurrence, study 1B offers additional evidence against a severity-based account of our findings.

Method

Six hundred and sixty UK residents were recruited on Prolific Academic. Five hundred and fifty-seven of them indicated that they had considered buying dietary supplements in the past year. These participants were shown an ad for Ginkgo Biloba—a dietary supplement sold on amazon.uk (web appendix 1)—and were randomly assigned to one of four conditions in a 2 (side effect: single vs. combined) × 2 (high risk: upset stomach vs. headaches) between-subjects design. Those in the single side effect condition learned that consuming Ginkgo Biloba can cause upset stomach (headaches), which happens in 45% of the people taking the supplement. Those in the combined high–low condition read that in addition to the 45% chance of upset stomach (headaches), there was a 5% chance of experiencing headaches (upset stomach). After excluding participants who did not pass the attention check, we were left with 512 participants (118 men, 387 women, 5 preferred not to say, and 2 identified as other; $M_{age} = 34.9$, $SD_{age} = 11.9$).

Participants rated the overall risk of experiencing side effects when taking the supplement on a 10-point sliding scale (0 = *very low risk*, 10 = *very high risk*) and indicated whether they would like to be entered in a lottery for a chance to receive a free bottle of the Ginkgo Biloba supplement (valued at £14.95). They also rated the severity of each side effect on three 11-point items (1 = *not a big deal*, 11 = *a huge deal*; 1 = *not bad at all*, 11 = *very bad*; 1 = *mild*, 11 = *severe*), rated their experience with making supplement-related decisions on a seven-point scale (1 = *no experience at all*, 7 = *a lot of experience*), and completed the Berlin numeracy test.

At the end of the study, participants were debriefed and asked to reconsent. The debrief explained the research hypothesis and the need to provide inaccurate risk information about the product used in the ad. Participants were informed that, in reality, the side effects described in this study were less frequent, while other side effects not described in the study were possible. They also read that the benefits described in the ad were those promoted by the marketer of the product and may or may not be scientifically supported. They were urged to always consult their physician before taking any dietary supplement. Finally, those who chose to be entered in a lottery for a chance to receive a free product were instead entered in a lottery to receive a £15 bonus, which was added to the lottery winner's Prolific account.

Results and Discussion

Perceived Risk. A two-way ANCOVA on perceived risk, using numeracy and experience as control variables, found a significant main effect of side effect ($F(1, 506) = 46.33$, $p < .001$, $\eta_p^2 = 0.084$). Neither the main effect of high risk ($F(1, 506) = 0.02$, $p = .88$, $\eta_p^2 = 0.000$) nor the side effect × high risk interaction ($F(1, 506) = 1.25$, $p = .26$, $\eta_p^2 = 0.002$) was significant. Consumers who read that the supplement comes with a high probability of experiencing an upset stomach and a low probability of experiencing headaches reported significantly lower overall risk ($M_{high–low} = 4.14$, $SD_{high–low} = 1.87$) than those who read about the high probability of upset stomach alone ($M_{high} = 5.54$, $SD_{high} = 2.14$; $t(506) = -5.62$, $p < .001$). Similarly, consumers who read that the supplement comes with a high probability of experiencing headaches and a low probability of experiencing upset stomach reported significantly lower overall risk ($M_{high–low} = 4.40$, $SD_{high–low} = 2.07$) than those who read about the high probability of headaches alone ($M_{high} = 5.38$, $SD_{high} = 1.99$; $t(506) = -4.02$, $p < .001$; figure 3). As for the covariates, numeracy had a positive effect on perceived risk ($b = 0.23$, SE = 0.08; $F(1, 506) = 7.50$, $p < .01$, $\eta_p^2 = 0.015$), whereas experience did not have significant effect ($b = 0.03$, SE = 0.05; $F(1, 253) = 0.37$, $p = .54$, $\eta_p^2 = 0.001$).

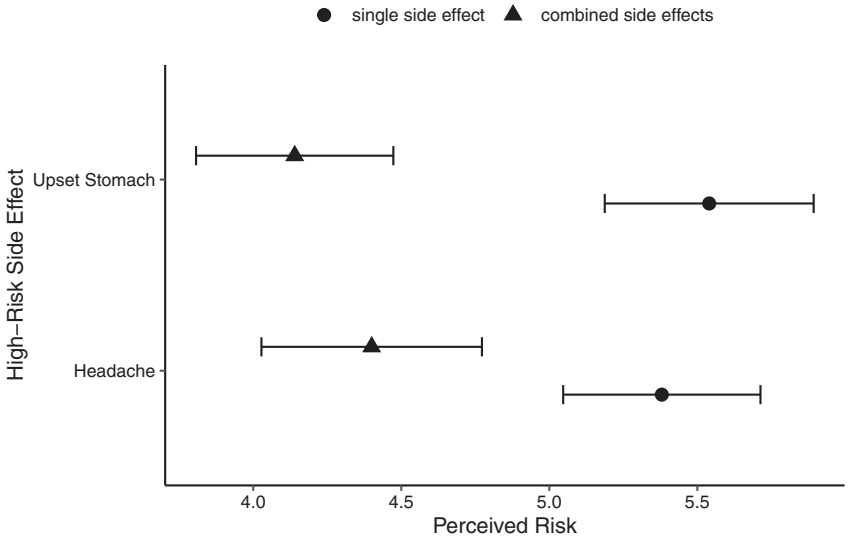
Choice to Enter the Lottery. When the high-risk side effect was upset stomach, participants who saw two side effects (52.0%) were more likely to enter the lottery for a free bottle of Ginkgo Biloba than those who saw a single side effect (38.1%, $\chi^2(1) = 4.53$, $p = .03$). This pattern was repeated when the high-risk side effect was headaches (figure 4): participants who saw two side effects (51.2%) were more likely to enter the lottery than those who saw a single side effect (36.2%, $\chi^2(1) = 5.27$, $p = .02$). Logistic regression analysis confirmed that the effect of single versus combined side effects on the log odds of entering the lottery does not depend on which side effect comes with a high probability of occurrence. Indeed, the side effect × high risk interaction was not significant ($b = 0.04$, SE = 0.36; $z = 0.11$, $p = .91$, OR = 1.04 [0.51, 2.12]).

Mediation Analysis. A mediation analysis using bias-corrected confidence intervals based on 5,000 bootstrap iterations revealed a non-significant direct effect of side effects on choice of entering the lottery (ADE = −0.05, 95% CI [−0.14, 0.03]), but a significant indirect effect through perceived risk (ACME = −0.09, 95% CI [−0.13, −0.06]). The direction of the indirect effect confirmed that consumers who saw the combined side effects perceived lower overall risk than those who saw only one side effect, which, in turn, contributed to increasing their likelihood to enter the lottery.

Perceived Severity. We tested for differences in the perceived severity of headaches and upset stomach. We

FIGURE 3

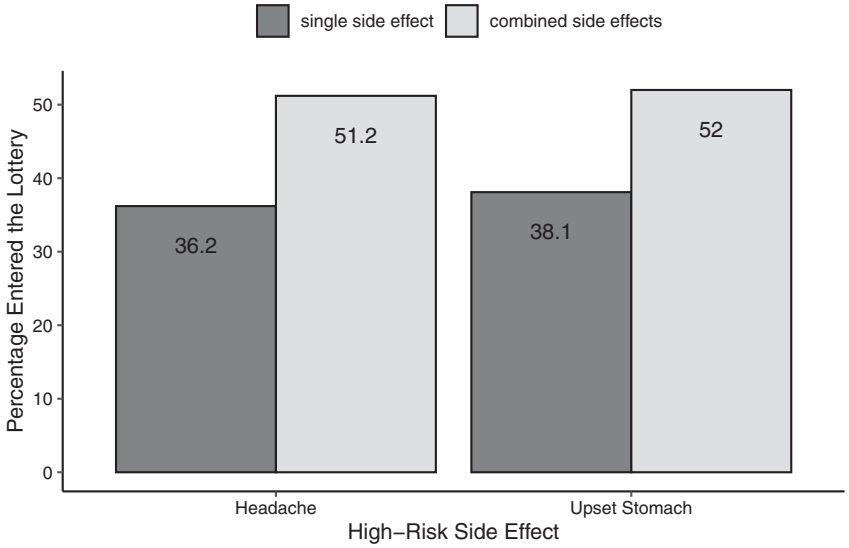
STUDY 1B: THE IMPACT OF SIDE EFFECTS TYPE AND NUMBER ON CONSUMERS' PERCEIVED RISK



NOTE.—Bars in graph represent 95% confidence intervals.

FIGURE 4

STUDY 1B: THE IMPACT OF SIDE EFFECTS TYPE AND NUMBER ON CONSUMERS' CHOICE OF ENTERING THE LOTTERY



created a severity index for each side effect by averaging participants' scores across the three severity items ($\alpha_{upset\ stomach} = 0.91$ and $\alpha_{headaches} = 0.91$). A paired samples *t*-test focused on those who saw both side effects found that headaches were considered more severe ($M = 7.18$, $SD = 2.09$) than abdominal cramps ($M = 6.35$, $SD = 2.25$; $t(247) = 5.82$, $p < .001$, $d = 0.37$ [0.24, 0.50]). A severity-based dilution account would predict that subtractive risk judgments should occur only when headaches are the primary (high-risk) side effect and upset stomach is the added, less severe side effect. Clearly, this prediction is incompatible with the data, since subtractive risk judgments occurred regardless of whether the SSE was headaches or upset stomach.

STUDY 1C: REPLICATION WITH MORE THAN TWO SIDE EFFECTS

In this study, we replicate previous studies' findings using yet another product. We also test whether categorical averaging occurs when there are more than two side effects.

Method

Five hundred US residents were recruited on Amazon MTurk. Four hundred and sixty of them indicated that they had considered buying health and wellness products in the past year. These participants were shown an ad for RENPHO—an eye massager sold on amazon.com (web appendix 1)—and were randomly assigned to one of three risk conditions (high vs. high–low vs. high–low–low). Those in the single high-risk condition learned that using RENPHO can cause local erythema (redness) and irritation of the eye lids, which happens in 40% of the people using the eye massager. Those in the combined high–low condition read that in addition to the 40% chance of redness and irritation, there was a 2% chance of experiencing dizziness. Finally, those in the combined high–low–low condition were informed that in addition to the 40% chance of redness and irritation and the 2% chance of dizziness, there was also a 1% chance of experiencing blurry vision. After excluding participants who did not pass the attention check, we were left with 382 participants (151 men, 227 women, 3 preferred not to say, and 1 identified as other; $M_{age} = 40.6$, $SD_{age} = 12.2$).

As in the previous study, participants rated the overall risk of experiencing side effects when using the eye massager on a 10-point sliding scale (0 = *very low risk*, 10 = *very high risk*) and indicated whether they would like to be entered in a lottery for a chance to receive a free RENPHO eye massager (valued at $69.99). They also rated their experience with making decisions about health and wellness products and completed the Berlin numeracy test. Finally, they were debriefed and asked to reconsent. The lottery winner received a $70 bonus.

Results and Discussion

Perceived Risk. A one-way ANCOVA on perceived risk, using numeracy and experience as control variables, found a significant effect of side effect condition ($F(2, 377) = 12.38$, $p < .001$, $\eta_p^2 = 0.062$). Pairwise comparisons with Bonferroni correction revealed that consumers who read about the combination of one high-risk side effect and a one low-risk side effect ($M_{high–low} = 4.40$, $SD_{high–low} = 2.15$) reported lower overall risk than those who read about a single high-risk side effect ($M_{single} = 5.35$, $SD_{single} = 2.00$; $t(377) = -3.49$, $p < .01$). Consumers who read about the combination of one high-risk side effect and two low-risk side effects also perceived lower overall risk than those who read about the single high-risk side effect ($M_{high–low–low} = 4.02$, $SD_{high–low–low} = 2.20$; $t(377) = -4.83$, $p < .001$). The difference between those who saw two side effects and those who saw three side effects was not significant ($t(377) = -1.38$, $p = .51$). In addition, neither numeracy ($b = 0.04$, $SE = 0.10$; $F(1, 377) = 0.20$, $p = .66$, $\eta_p^2 = 0.001$) nor experience ($b = 0.13$, $SE = 0.14$; $F(1, 377) = 0.89$, $p = .35$, $\eta_p^2 = 0.002$) had a significant effect on perceived risk.

Choice to Enter the Lottery. Participants who saw two side effects (66.4%) or three side effects (69.5%) were more likely to enter the lottery for a free eye massager than those who saw a single side effect (53.2%). Logistic regression analysis showed that these differences were statistically significant ($b_{high–low} = 0.60$, $SE_{high–low} = 0.26$; $z = 2.29$, $p = .02$, $OR = 1.83$ [1.10, 3.07], and $b_{high–low–low} = 0.77$, $SE_{high–low–low} = 0.27$; $z = 2.87$, $p < .01$, $OR = 2.15$ [1.28, 3.66]). Experience with making decisions about health and wellness products was positively associated with likelihood to enter the lottery ($b = 0.31$, $SE = 0.17$; $z = 2.35$, $p = .02$, $OR = 1.49$ [1.09, 2.14]). However, numeracy was not ($b = 0.09$, $SE = 0.10$; $z = 0.78$, $p = .89$, $OR = 1.08$ [0.89, 1.32]).

Mediation Analysis. We conducted a mediation analysis, examining the bias-corrected confidence intervals using 5,000 bootstrap iterations. When contrasting the high–low with the high-only conditions, the direct effect of side effect condition on choice of entering the lottery (ADE = 0.10, 95% CI [−0.02, 0.22]) was not significant, but the indirect effect through perceived risk (ACME = 0.04, 95% CI [0.01, 0.08]) was significant. The direction of the indirect effect confirmed that consumers who saw two side effects perceived lower overall risk than those who saw only one side effect, which, in turn, contributed to increasing their likelihood to enter the lottery. Contrasting the high–low–low condition with the high-only condition yielded similar results. Side effect condition had a non-

significant direct effect on choice of entering the lottery (Average Direct Effect [ADE] = 0.12, 95% CI [−0.0002, 0.24]), and a significant indirect effect through perceived risk (Average Causal Mediation Effect [ACME] = 0.05, 95% CI [0.02, 0.10]).

We proposed that when combining items with qualitatively different risk levels, consumers tend to average across the categorical judgments of high risk and low risk instead of adding the numeric risks, which leads to an overall risk assessment for the combined items that is lower than the perceived risk associated with the high-risk item alone. Study 1 (A, B, C) provides robust evidence that consumers indeed judge combined items with high- and low-risk levels to be less risky overall than the high-risk item alone and that this difference in perceived risk, in turn, affects their behavior in consequential tasks. However, study 1 does not test the underlying process.

STUDY 2: PROBABILITIES IN CONTEXT

Categorical averaging implies that consumers do not average the numerical probabilities they encounter, but the cruder risk categories that these probabilities represent. This study offers a direct test of this implication.

Categorical representations of numerical probabilities are sensitive to the context in which the probabilities are encountered. A given numerical probability (e.g., 20%) would be classified in a higher risk category when encountered in the context of lower probabilities (e.g., 1% or 2%) than when encountered in the context of higher probabilities (e.g., 60% or 70%). Since averaging is predicted to only occur when combining items in different risk tiers, we expect to see subtractive risk judgments only when an added side effect with a given numerical probability is categorized as a lower risk.

Method

Three hundred and fifty US residents recruited from Prolific Academic participated in study 2 in exchange for monetary compensation. They were randomly assigned to one of three conditions (single vs. combined high vs. combined low). After excluding participants who did not pass the attention check, we were left with 313 participants (166 men, 137 women, 6 preferred not to say, and 4 identified as other; $M_{age} = 32.5$, $SD_{age} = 10.3$).

Participants were presented with the arthritis medical scenario from study 1A with modified risk values. Those in the single side effect condition learned that the medication comes with a 55% chance of experiencing blurry vision and increased light sensitivity. Those in the other two conditions learned that in addition to the 55% chance of blurry vision and increased light sensitivity, there was a 20% chance of experiencing abdominal cramps and pain.

To manipulate the categorization of the 20% numerical probability as either high or low risk, participants in the combined high and combined low conditions were asked to review four brands of sleeping pills prior to viewing the arthritis scenario. The four brands of sleeping pills varied in their probability of causing abdominal cramps (web appendix 2). One brand (drug A) was described as having a 20% probability of causing abdominal cramps and pain and was presented in conjunction with three other brands described as having either significantly lower or significantly higher probabilities of causing the same side effect. In the combined high condition, the probabilities for drugs B, C, and D were 4%, 2%, and 1%, respectively, thus making a 20% probability of abdominal cramps appear as a high risk. In the combined low condition, the probabilities for drugs B, C, and D were 55%, 62%, and 69%, respectively, making a 20% probability of abdominal cramps appear as a low risk. After reviewing the four brands of sleeping pills, participants classified each drug's risk of causing abdominal cramps and pain as either high or low.

All participants rated their perceived overall risk of side effects associated with the arthritis medication on a 10-point scale (0 = *very low risk*, 10 = *very high risk*), their likelihood to start the medication on a 10-point scale (0 = *not at all likely*, 10 = *very likely*), their experience with making medication-related decisions on a 7-point scale (1 = *no experience at all*, 7 = *a lot of experience*), and took the Berlin numeracy test.

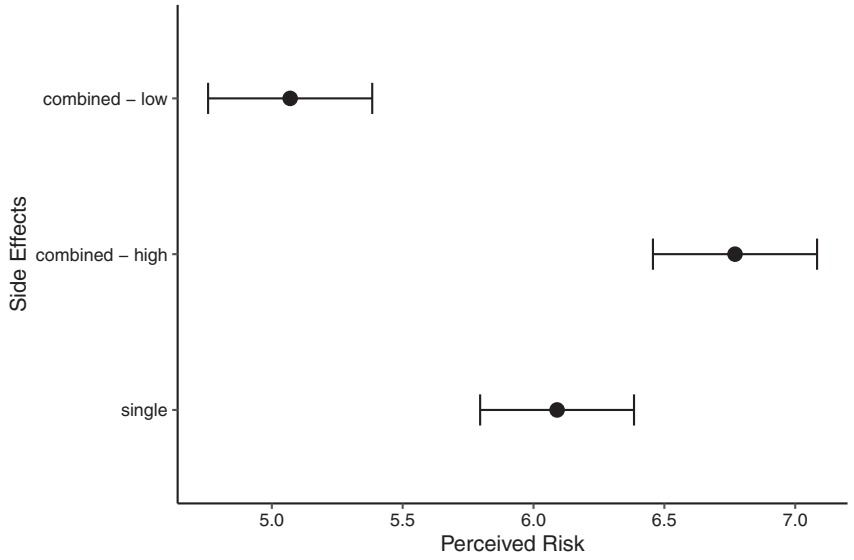
Results and Discussion

Manipulation Check. As expected, participants in the combined high condition were more likely to classify the 20% probability of experiencing abdominal cramps as high risk than low risk (65.1% vs. 34.9%, $\chi^2(1) = 9.66$, $p = .002$). Conversely, participants in the combined low condition were more likely to classify the same 20% probability as low risk than high risk (71.8% vs. 28.2%, $\chi^2(1) = 19.66$, $p < .001$).

Perceived Risk. A one-way ANCOVA on perceived overall risk, using numeracy and experience with making medication-related decisions as control variables, found a significant effect of side effect condition ($F(2, 308) = 29.97$, $p < .001$, $\eta_p^2 = 0.16$). Planned contrasts with Bonferroni correction for multiple tests revealed that consumers in the combined low condition ($M = 5.07$, $SD = 1.66$) reported lower overall risk than those who read about a single high-risk side effect ($M = 6.09$, $SD = 1.49$; $t(308) = -4.65$, $p < .001$). In contrast, consumers in the combined high condition ($M = 6.77$, $SD = 1.60$) perceived greater overall risk than those in the single side effect condition ($M = 6.09$, $SD = 1.49$; $t(308) = 2.99$, $p = .006$). Furthermore, neither numeracy ($b = 0.04$, $SE = 0.09$; $F(1, 308) = 0.20$, $p = .65$, $\eta_p^2 = 0.001$) nor experience

FIGURE 5

STUDY 2: THE IMPACT OF SIDE EFFECTS ON PERCEIVED RISK



NOTE.—Bars in graph represent 95% confidence intervals.

($b = -0.02$, SE = 0.05; $F(1, 308) = 0.19$, $p = .66$, $\eta_p^2 = 0.001$) had a significant effect on perceived risk. These results are consistent with the notion that consumers do not average numerical probabilities but the risk categories that these probabilities represent. Figure 5 illustrates the influence of side effects on perceived risk.

Likelihood to Start. A one-way ANCOVA on likelihood to start, using numeracy and experience with making medication-related decisions as control variables, found a significant effect of side effect condition ($F(2, 308) = 14.92$, $p < .001$, $\eta_p^2 = 0.09$). Planned contrasts with Bonferroni correction for multiple tests revealed that consumers in the combined low condition ($M = 6.66$, SD = 1.97) reported higher likelihood to start than those who read about a single high-risk side effect ($M = 5.76$, SD = 2.44; $t(308) = 2.83$, $p = .01$). In contrast, consumers in the combined high condition ($M = 4.90$, SD = 2.54) reported lower likelihood to start the medication than those in the single side effect condition ($M = 6.09$, SD = 1.49; $t(308) = -2.60$, $p = .02$). Furthermore, neither numeracy ($b = 0.08$, SE = 0.13; $F(1, 308) = 0.37$, $p = .54$, $\eta_p^2 = 0.001$) nor experience ($b = 0.13$, SE = 0.08; $F(1, 308) = 3.05$, $p = .08$, $\eta_p^2 = 0.010$) had a significant effect on likelihood to start. Figure 6 illustrates the influence of side effects on likelihood to start.

Mediation Analysis. We conducted a mediation analysis, focusing on the combined low and single side effect conditions. We examined the bias-corrected confidence intervals using 5,000 bootstrap iterations. The direct effect of side effect condition on likelihood to start the medication was not significant (ADE = 0.37, 95% CI [−0.21, 0.95]), but the indirect effect through perceived risk was significant (ACME = 0.58, 95% CI [0.31, 0.91]). The direction of the indirect effect confirmed that consumers who saw the combined side effects perceived lower overall risk than those who saw only one side effect, which, in turn, contributed to increasing their likelihood to start the medication.

STUDY 3: WHEN THE RISK IS TOO HIGH

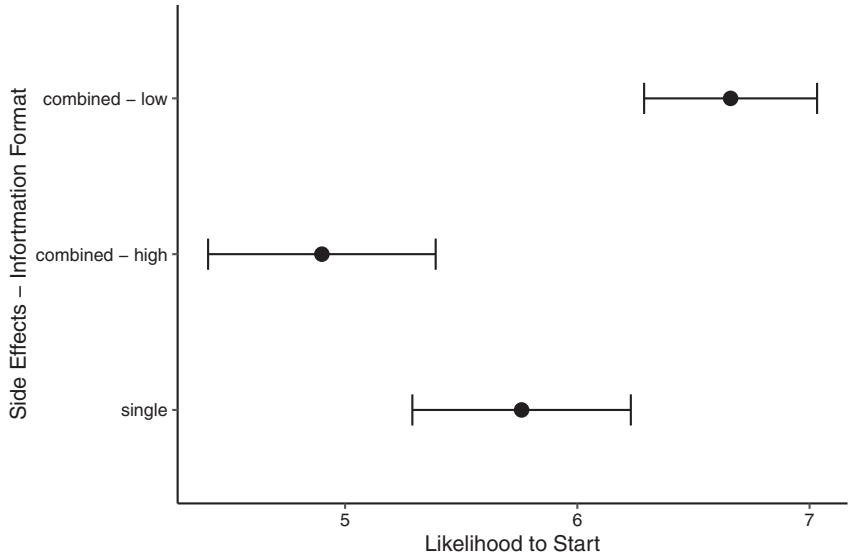
This study tests the notion that categorical averaging is less likely when the risk of the FSE is too high. We reasoned that the risk of a highly probable side effect may disproportionally account for consumers' judgment of overall risk. This study also explores the robustness and boundaries of categorical averaging across a wide range of probabilities for the SSE.

Method

One thousand six hundred US residents recruited from Prolific Academic completed study 3 in exchange for

FIGURE 6

STUDY 2: THE IMPACT OF SIDE EFFECTS ON LIKELIHOOD TO START



NOTE.—Bars in graph represent 95% confidence intervals.

monetary compensation. They were randomly assigned to one of 14 conditions in a 2 (FSE: 45% vs. 75%) × 7 (SSE: none vs. 1% vs. 6% vs. 17% vs. 22% vs. 30% vs. 35%) between-subjects design. After excluding participants who did not pass the attention check, we were left with 1,551 participants (772 men, 726 women, 9 preferred not to say, and 44 identified as other; $M_{age} = 35.1$, $SD_{age} = 13.0$).

Participants responded to a different medical scenario adapted from Promberger and Baron (2006), in which they imagined having trouble going to sleep (web appendix 1). Those in the single side effect conditions read that the new medication can cause headaches for either 45% (high) or 75% (too high)[2] of the people who take it. Those in the combined side effects condition learned that in addition to the chance of headaches, there was either a 1%, 6%, 17%, 22%, 30%, or 35% chance of experiencing heartburns. All participants rated the overall risk of side effects, likelihood to start the medication, experience with making medication-related decisions and completed the Berlin numeracy test.

2 The pilot study revealed that of the 10 tested probabilities of side effects, 45% was the most represented in the "high" category, while 75% was the most represented in the "too high" category. Importantly, participants rated the 45% probability as "high" more frequently than they rated it as "too high" (115 vs. 16, $\chi^2(1) = 74.8$, $p < .001$). Conversely, they rated the 75% probability as "too high" more frequently than they rated it as "high" (179 vs. 26, $\chi^2(1) = 114.2$, $p < .001$).

Results and Discussion

A two-way ANCOVA on perceived risk, using numeracy and experience with making medication-related decisions as control variables, found a significant main effect of the FSE condition ($F(1, 1535) = 62.38$, $p < .001$, $\eta_p^2 = 0.039$), a significant main effect of the SSE condition ($F(6, 1535) = 21.63$, $p < .001$, $\eta_p^2 = 0.078$), and a significant interaction between the FSE and SSE conditions ($F(6, 1535) = 8.85$, $p < .001$, $\eta_p^2 = 0.033$). Neither experience with making medication decisions ($b = 0.02$, SE = 0.03; $F(1, 1535) = 1.67$, $p = .20$, $\eta_p^2 = 0.001$) nor numeracy ($b = -0.01$, SE = 0.04; $F(1, 1535) = 0.02$, $p = .88$, $\eta_p^2 < 0.001$) had a significant effect on perceived risk.

As expected, when the probability of headaches was set to 75%, overall risk perceptions of consumers who saw a single side effect were high ($M_{FSE75} = 7.55$, $SD_{FSE75} = 1.23$) and did not differ significantly from those who saw two side effects, regardless of the SSE's probability of occurrence (table 1).

In contrast, when the probability of headaches was set at 45%, consumers who read about the possibility of experiencing both headaches and heartburns reported significantly lower overall risk than those who read about only the possibility of headaches ($M_{FSE45} = 5.70$, $SD_{FSE45} = 1.83$) when the probability of heartburns was either 1% ($M_{FSE45\_SSE1} = 4.69$, $SD_{FSE45\_SSE1} = 1.77$; $t(1535) = -4.24$, $p < .001$), 6% ($M_{FSE45\_SSE6} = 4.57$, SD_{FSE45_SSE6}

TABLE 1

PERCEIVED OVERALL RISK ACROSS CONDITIONS

Condition	*N*	Mean	SD	Mean difference	*t*-Ratio (*df* = 1535)	*p*-Value
FSE45	110	5.70	1.83			
FSE45_SSE1	105	4.69	1.77	−1.01	−4.27	<.001
FSE45_SSE6	111	4.57	1.63	−1.13	−4.85	<.001
FSE45_SSE17	112	4.89	1.69	−0.81	−3.45	.003
FSE45_SSE22	112	5.15	1.51	−0.55	−2.35	.113
FSE45_SSE30	110	6.37	1.97	0.67	2.88	.025
FSE45_SSE35	106	6.43	1.96	0.73	3.07	.013
FSE75	112	7.55	1.23			
FSE75_SSE1	110	7.28	1.84	−0.27	−1.16	>.999
FSE75_SSE6	112	7.51	1.46	−0.03	−0.13	>.999
FSE75_SSE17	112	7.19	1.84	−0.36	−1.52	.780
FSE75_SSE22	115	7.29	1.95	−0.26	−1.12	>.999
FSE75_SSE30	113	7.48	1.79	−0.07	−0.29	>.999
FSE75_SSE35	111	7.57	1.72	0.02	0.12	>.999

NOTE.—Mean differences and associated statistical tests contrast each combined side effects condition with its corresponding single side effect condition. *p*-Values are adjusted using Bonferroni correction for multiple tests.

= 1.63; $t(1535) = -4.85$, $p < .001$), or 17% ($M_{FSE45\_SSE17} = 4.89$, $SD_{FSE45\_SSE17} = 1.69$; $t(1535) = -3.45$, $p = .003$). Overall perceived risk was also lower when the probability of heartburn was 22% ($M_{FSE45\_SSE22} = 5.15$, $SD_{FSE45\_SSE22} = 1.51$; $t(1535) = -2.35$, $p = .11$), though the difference was not statistically significant. However, when the probability of heartburns was either 30% ($M_{FSE45\_SSE30} = 6.37$, $SD_{FSE45\_SSE30} = 1.97$; $t(1535) = 2.88$, $p = .03$) or 35% ($M_{FSE45\_SSE35} = 6.43$, $SD_{FSE45\_SSE35} = 1.96$; $t(1535) = 3.07$, $p = .01$), those who read about both side effects reported greater overall risk than those who read about only headaches. These results—summarized in table 1 and illustrated in figure 7—offer interesting insights on the boundaries of categorical averaging. First, they suggest that categorical averaging is less likely when the risk of the FSE is exceedingly high. In this situation, impressions of overall risk appear to be driven primarily by perceptions of the FSE. Second, they confirm that even when the probability of the FSE is not exceedingly high, categorical averaging does not happen when the risk of the SSE is also high. The results also suggest that an important shift in the categorization of risk happens at a probability value somewhere between 22% and 30%. This is consistent with the findings from the pilot study. However, as discussed earlier, the mapping of numerical probabilities onto risk categories is likely to be person and context dependent.

In study 3, a side effect's riskiness was manipulated by varying its numerical probability. However, risk perception is a function of both probability and severity. Thus, we reasoned that a low-probability side effect that is extremely harmful would be categorized as a high-risk side effect. To examine the impact of adding an extremely harmful side effect, we conducted an additional study (web appendix 9) comparing the overall risk perceptions in four conditions: a single high-risk side effect, a combination of one high-risk and one low-risk side effect, a combination of two high-risk side effects, and a combination of one high risk and one low risk but extremely severe side effects. As in all the previous studies, adding a low-risk side effect led to subtractive risk judgments. However, adding a low-risk side effect that was extremely harmful had the same effect as adding a high-risk side effect. In both cases, the overall perceived risk was greater than in the case of a single high-risk side effect.

STUDY 4: ASSESSMENT OF RISK VERSUS PROBABILITY

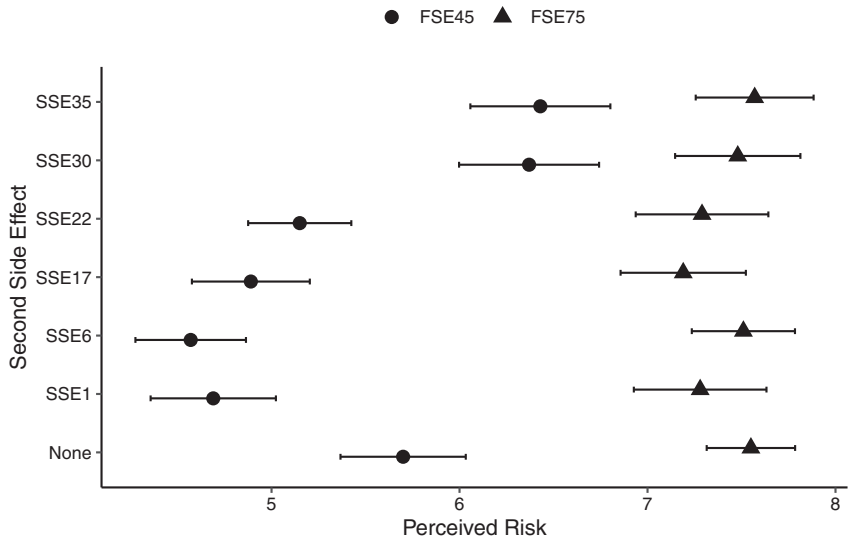
This study tests the impact of asking about probabilities versus risk. Averaging occurs because consumers think categorically about risk. When estimating the overall risk of side effects, they tend to rely on the gist representations of the numerical probabilities to form a categorical impression of the overall risk. Averaging, however, is unlikely when consumers are explicitly tasked with estimating numerical probabilities. In this situation, they are more likely to rely on the verbatim representation of the numbers which map directly onto the response scale. Thus, we expect that when asked to estimate the probability of experiencing any side effect, consumers will report higher probabilities for two side effects than for a single side effect. Moreover, if consumers are indeed reporting their categorical impressions when asked to estimate the overall risk of side effects, then we would expect their answers to follow the same pattern when responding on a more categorical scale.

Method

Eight hundred US residents recruited from Prolific participated in this study in exchange for monetary

FIGURE 7

STUDY 3: CATEGORICAL AVERAGING ACROSS A WIDE RANGE OF PROBABILITIES



NOTE.—Bars in graph represent 95% confidence intervals.

compensation. They were randomly assigned to one of six conditions in a 2 (side effect: single vs. combined) × 3 (judgment task: risk vs. risk cat vs. probability) between-subjects design. After excluding participants who did not pass the attention check, we were left with 783 participants (394 men, 376 women, 6 preferred not to say, and 7 identified as other; $M_{age} = 34.4$, $SD_{age} = 12.3$).

Participants read about the sleep medication scenario from study 3. Those in the single side effect condition read that a new medication that may help them can cause headaches for 45% of the people who take it. Those in the combined side effects condition learned that in addition to the 45% chance of headaches, there was a 6% chance of experiencing heartburn.

Participants in the risk judgment task condition rated their risk perceptions on a 100-point sliding scale anchored by *very low risk* and *very high risk* as in the previous studies. Those in the risk-cat condition rated their risk perceptions on an augmented 100-point sliding scale that included five risk labels (*very low*, *low*, *moderate*, *high*, and *very high*) to enhance the categorical perception of the response. We expect no significant difference in perceived risk across the risk and risk-cat conditions, as participants in both conditions think categorically when assessing the overall risk. Participants in the probability condition were asked to estimate the probability of experiencing any side effect on a 100-point sliding scale anchored by 0% and 100%. All participants indicated their likelihood to start the medication on a 100-point sliding scale (0 = *not likely at all*, 100 = *very likely*), took the Berlin numeracy test, answered demographic questions, and reported on their experience with making medication-related decisions.

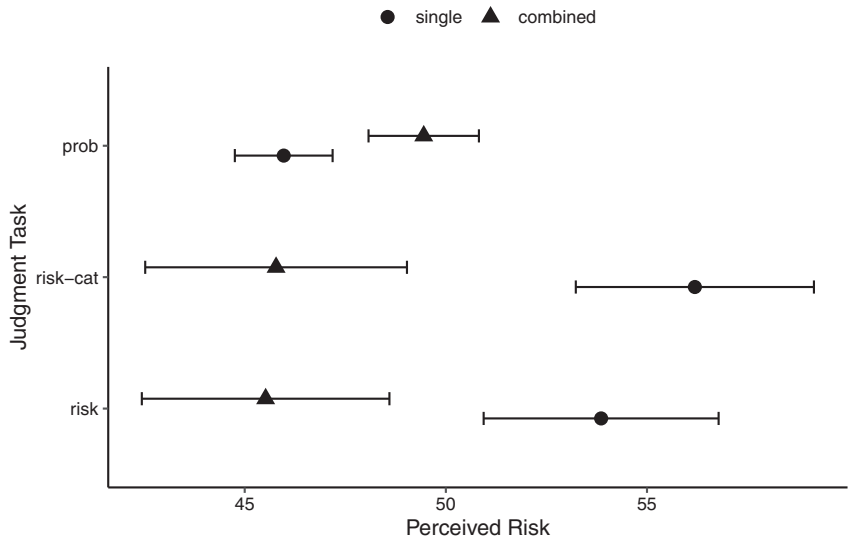
Results and Discussion

A two-way ANCOVA on perceived risk, using numeracy and experience with making medication-related decisions as control variables, found a significant effect of side effect ($F(1, 775) = 20.24$, $p < .001$, $\eta_p^2 = 0.025$), a significant effect of judgment task ($F(2, 775) = 16.47$, $p < .001$, $\eta_p^2 = 0.041$), and a significant interaction between side effect and judgment task ($F(2, 775) = 16.13$, $p < .001$, $\eta_p^2 = 0.040$). Moreover, numeracy ($b = -0.45$, SE = 0.50; $F(1, 775) = 0.82$, $p = .36$, $\eta_p^2 = 0.001$) and experience with making medication decisions ($b = 0.60$, SE = 0.32; $F(1, 775) = 3.43$, $p = .06$, $\eta_p^2 = 0.004$) did not significantly influence perceived risk.

When consumers were asked to estimate the overall risk of side effects using a scale anchored by *very low risk* and *very high risk*, as in the previous studies, those who read about two side effects ($M = 45.5$, $SD = 17.9$) reported lower overall risk than those who read about a single side effect ($M = 53.9$, $SD = 16.9$; $t(775) = -4.50$, $p < .001$). Subtractive risk judgments were also observed when the risk scale included more labels. Those who read about both

FIGURE 8

STUDY 4: EFFECTS OF SIDE EFFECTS AND JUDGMENT TASK ON PERCEIVED RISK



NOTE.—Bars in graph represent 95% confidence intervals.

side effects ($M = 45.8$, SD = 18.8) reported lower overall risk than those who read about a single side effect ($M = 56.2$, SD = 17.3; $t(775) = -5.51$, $p < .001$). In contrast, when consumers were asked to estimate the probability of any side effect on a scale anchored by 0% and 100%, those who read about the possibility of two side effects ($M = 49.4$, SD = 8.09) correctly reported a higher probability than those who read about a single side effect, and the difference was marginally significant ($M = 46.0$, SD = 7.1; $t(775) = 1.87$, $p = .06$). These results are illustrated in figure 8.

We also examined how the manipulation of the judgment task influenced risk perception in each side effect condition. When consumers read about a single side effect, their overall risk perception did not differ significantly between the risk and risk-cat judgment conditions ($t(775) = -1.13$, $p = .77$). However, those in the probability judgment condition provided lower estimates than both those in the risk judgment condition ($t(775) = -4.28$, $p < .001$) and those in the risk-cat judgment condition ($t(775) = -5.44$, $p < .001$). These patterns reflect the fact that, in the probability judgment condition, the estimates were more in line with the numeric probabilities provided in the scenario.

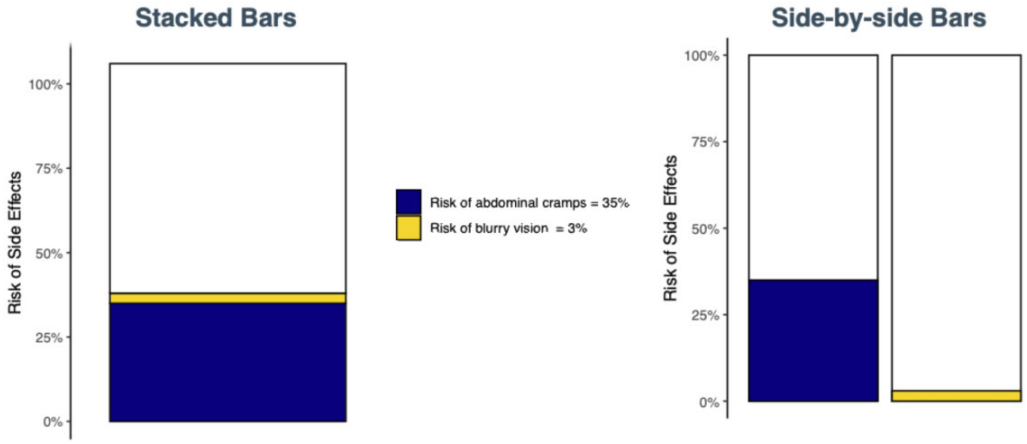
When participants read about two side effects, those in the probability judgment condition reported slightly higher risk scores than those in the risk judgment condition ($t(775) = 2.10$, $p = .11$) and those in the risk-cat judgment conditions ($t(775) = 1.96$, $p = .15$), though the differences were not statistically significant. Moreover, participants in the latter two conditions reported almost identical risk scores ($t(775) = 0.13$, $p > .99$). Here too, overall risk perceptions in the probability judgment condition were highly anchored on the numeric probabilities provided in the scenario. These patterns are informative but the relevant contrasts for testing our hypothesis are those comparing risk perceptions of single versus combined side effects within each judgment task. As noted earlier, subtractive risk judgments were found in the risk and risk-cat judgment conditions but not in the probability judgment condition. In the latter, consumers correctly judged the medication with two side effects to be risker than the medication with only one side effect.

STUDY 5A: GRAPHICAL PRESENTATION OF RISK INFORMATION: BAR GRAPHS

In this study, we test the effectiveness of presenting risk information in a graphical format. We reasoned that graphical presentations that highlight the additive nature of multiple risks could reduce categorical averaging and lower consumers' biased perceptions of multiple risks. Moreover, this effect should be specific to graphs that highlight the

FIGURE 9

STUDY 5A: GRAPHICAL PRESENTATION OF RISK USING STACKED VERSUS SIDE-BY-SIDE BARS



additive nature of the multiple risks, not just any graphical presentation.

Method

Seven hundred and fifty participants from Prolific Academic took part in study 5A in exchange for monetary compensation. They were randomly assigned to one of five side effect/presentation format conditions (single side effect/numerical vs. single side effect/graph vs. combined side effects/numerical vs. combined side effects/graph emphasizing additive risks vs. combined side effects/graph not emphasizing additive risks). After excluding participants who failed the attention check, we were left with 664 respondents (387 men, 267 women, 3 preferred not to say, and 7 identified as other; $M_{age} = 31.4$, $SD_{age} = 11.1$).

Participants in the numerical presentation conditions responded to the arthritis medication scenarios from study 1A with slightly modified risk values (35% for abdominal cramps and 3% for blurry vision).

Participants in the graph conditions responded to the same scenario augmented by bar graphs. Stacked bars were used in the graph format emphasizing additive risks, and side-by-side bars were used in the graph format not emphasizing additive risks (figure 9). We hypothesized that because stacked bars help visualize the additive nature of multiple risks, they would evoke an additive integration rule, thus reducing categorical averaging and resulting in higher risk perceptions than when information is presented in a numerical format. Side-by-side bars, however, do not emphasize the additive property of multiple risks, and as a result are unlikely to reduce categorical averaging. Side-by-side bars were expected to result in the same subtractive judgments as in the numerical format.

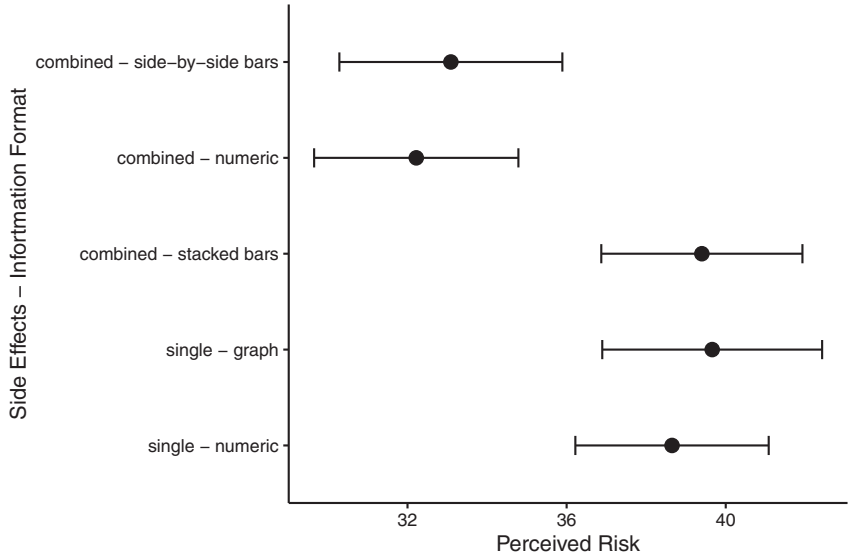
Participants rated the overall risk of side effects, likelihood to start the medication, and experience with making medication-related decisions as in study 1A. They also completed the abbreviated numeracy scale (Weller et al. 2013) instead of the Berlin numeracy scale. The abbreviated numeracy scale (web appendix 2) consists of eight open-ended items. The following is a sample item: "Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?" Answers are coded 1 for a correct answer and 0 for an incorrect answer. The final numeracy score is the sum of all correct answers.

Results and Discussion

A one-way ANCOVA on perceived risk, using side effect/presentation format as the independent variable and numeracy and experience with making medication-related decisions as control variables, revealed a significant effect of side effect/presentation format on perceived risk ($F(4, 657) = 6.68$, $p < .001$, $\eta_p^2 = 0.039$). We conducted six planned contrasts with Bonferroni correction for multiple comparisons. Consistent with the previous studies and with categorical averaging, when the risk information was presented only numerically, consumers who were exposed to the combination of one high-risk side effect and one low-risk side effect reported lower overall risk than those who were exposed to only the high-risk side effect ($M_{combined/numerical} = 32.2$, $SD = 15.2$ vs. $M_{single/numerical} = 38.6$, $SD = 14.3$; $t(657) = -3.21$, $p < .01$). Moreover, visualizing the combined risks using side-by-side bars did not

FIGURE 10

STUDY 5A: STACKED BARS REDUCE CATEGORICAL AVERAGING AND LEAD TO HIGHER RISK ESTIMATES THAN NUMERIC AND SIDE-BY-SIDE FORMATS



NOTE.—Bars in graph represent 95% confidence intervals.

decrease categorical averaging, as consumers in the side-by-side graph condition also reported lower overall risk than those in the single side effect condition ($M_{combined/sbs}$ = 33.1, SD = 16.7; $t(657) = -2.81$, $p = .03$). In contrast, presenting the combined risks using stacked bars resulted in less categorical averaging. Indeed, not only did participants in the stacked bars condition not perceive lower overall risk than those in the single side effect/numerical condition ($M_{combined/stacked} = 39.4$, SD = 14.5; $t(657) = 0.40$, $p = 1.00$), they also reported higher overall risk than those in the combined side effects/numerical condition ($t(657) = 3.57$, $p < .01$) and those in the combined side effects/side-by-side graph condition ($t(657) = 3.18$, $p < .01$). Perceived risk in the single side effect/graph condition ($M_{single/graph} = 39.7$, SD = 16.4) did not differ significantly from that in the single side effect/numerical condition ($t(657) = 0.54$, $p = 1.00$). Finally, with respect to the effects of the covariates, there was a significant negative effect of numeracy ($b = -0.87$, SE = 0.36; $F(1, 657) = 5.87$, $p = .02$, $\eta_p^2 = 0.009$); also, there was a nonsignificant effect of experience ($b = 0.07$, SE = 0.33; $F(1, 657) = 0.04$, $p = .84$, $\eta_p^2 = 0.000$). Figure 10 illustrates the differences in perceived risk across the five conditions.

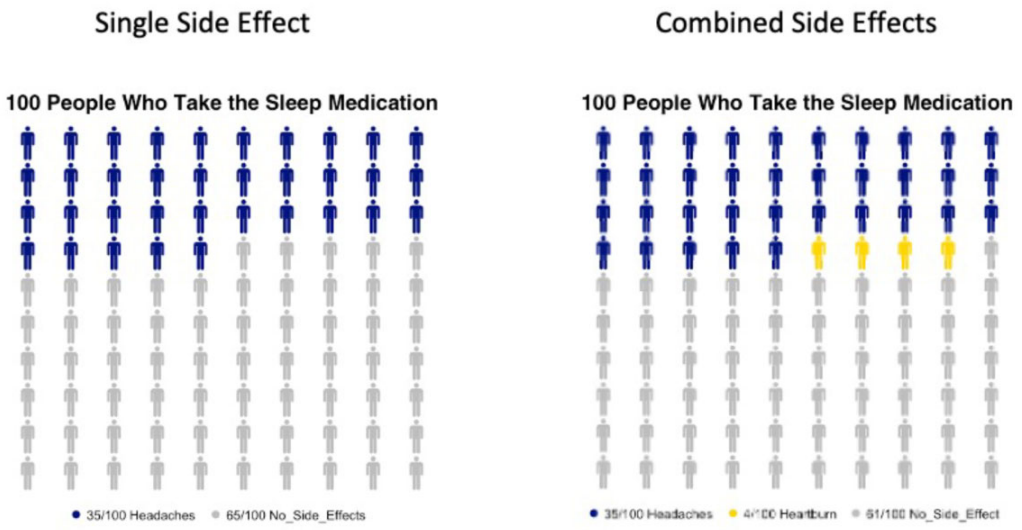
We reasoned that providing risk information in a graphical format that emphasizes the additive nature of multiple risks could be a simple and effective intervention aimed at reducing consumers' tendency to average across high and low risks. This study provides strong support for this idea. When risk information was presented numerically, those who saw a combination of one high risk and one low risk perceived lower overall risk than those who saw the high risk alone, in line with categorical averaging. However, when multiple risk information was presented graphically showing the low risk stacked on top of the high risk, consumers did not perceive the combined risks to be lower than the high risk alone. Importantly, this effect did not extend to graphical presentations that did not highlight the additive property of multiple risks. For example, presenting the high and low risks as side-by-side bars resulted in similar overall risk assessment to when the information was presented numerically. These results suggest that simply using graphs to communicate risk information is not a panacea for improving multiple risk perception. However, they also suggest that presentation formats other than stacked bars may be equally effective so long as they visually communicate the additive nature of multiple risks. In the next study, we test the effectiveness of pictographs as an alternative presentation format.

STUDY 5B: GRAPHICAL PRESENTATION OF RISK INFORMATION: PICTOGRAPHS

In this preregistered study (https://osf.io/bjmhv), we use a different medical scenario and examine the efficacy of

FIGURE 11

STUDY 5B: GRAPHICAL PRESENTATION OF RISK USING PICTOGRAPHS



pictographs as an alternative graphical intervention. While pictographs have been shown to improve people's understanding of medical risk (Trevena et al. 2013), their prior use has mainly focused on visualizing single risks. Pictographs can, however, be used to present information about multiple risks in a way that highlights their additive property. Thus, as with stacked bars, we expect pictographs to reduce categorical averaging and lower consumers' biased perceptions of multiple risks.

Method

Six hundred and sixty US residents recruited from Prolific completed this study in exchange for monetary compensation. They were randomly assigned to one of four conditions in a 2 (side effect: single vs. combined) × 2 (format: numeric vs. graph) between-subjects design. After excluding participants who did not pass the attention check, we were left with 652 participants (323 men, 318 women, 1 preferred not to say, and 10 identified as other; $M_{age} = 37.6$, $SD_{age} = 13.5$).

Participants responded to the sleep medication scenario from study 3A with modified risk values. Those in the single side effect condition read that the new medication can cause headaches for 35% of the people who take it. Those in the combined side effects condition learned that in addition to the 35% chance of headaches, there was a 4% chance of experiencing heartburn.

Participants in the graph condition responded to the same scenarios augmented by pictographs depicting an array of 100 people icons (figure 11). All participants rated the overall risk of side effects, likelihood to start the medication, experience with making medication-related decisions and completed the Berlin numeracy test as in the previous studies.

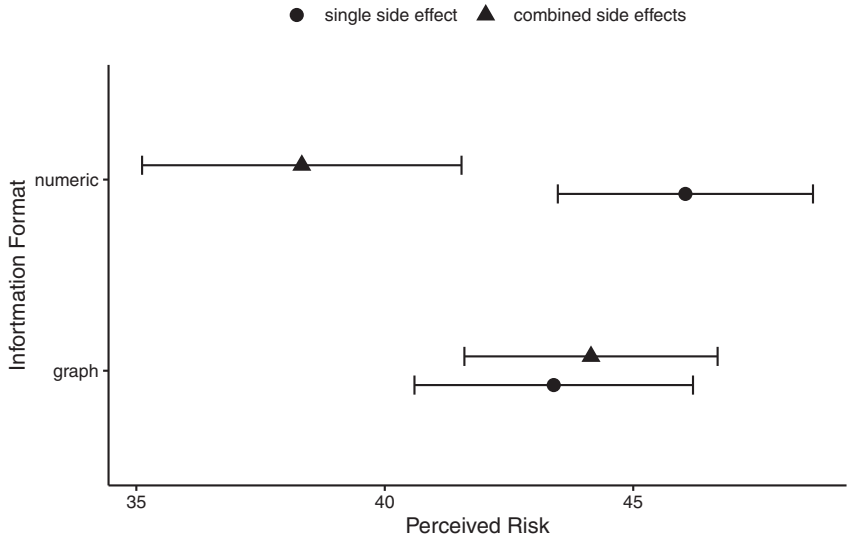
Results and Discussion

A two-way ANCOVA on perceived risk, using numeracy and experience with making medication-related decisions as control variables, found no main effect of side effect ($F(1, 646) = 0.74$, $p = .79$, $\eta_p^2 < 0.001$) or presentation format ($F(1, 646) = 1.67$, $p = .20$, $\eta_p^2 = 0.003$), but a significant interaction between side effect and presentation format ($F(1, 646) = 8.38$, $p < .001$, $\eta_p^2 = 0.013$). Moreover, experience with making medication decisions ($b = 0.13$, SE = 0.41; $F(1, 646) = 0.10$, $p = .75$, $\eta_p^2 < 0.001$) did not influence perceived risk, whereas numeracy ($b = -1.91$, SE = 0.65; $F(1, 646) = 8.59$, $p = .003$, $\eta_p^2 = 0.013$) had a negative effect on perceived risk.

When risk information was presented numerically, consumers who read about the combination of one high-risk side effect and one low-risk side effect ($M_{combined} = 38.3$, $SD_{combined} = 20.9$) reported lower overall risk than those who read about a single high-risk side effect ($M_{single} = 46.0$, $SD_{single} = 16.8$; $t(646) = -7.71$, $p < .001$). In contrast, when risk information was presented graphically, consumers who saw a single side effect ($M_{single} = 43.4$, $SD_{single} = 18.2$) reported overall risk that did not differ significantly from that reported by consumers who saw two

FIGURE 12

STUDY 5B: PICTOGRAPHS REDUCE CATEGORICAL AVERAGING AND LEAD TO HIGHER RISK ESTIMATES THAN NUMERIC FORMAT



NOTE.—Bars in graph represent 95% confidence intervals.

side effects ($M_{combined} = 44.2$, $SD_{combined} = 16.6$; $t(646) = 0.27$, $p = .79$; figure 12).

We also examined how the presentation format influenced risk perception in each side effect condition. When consumers read about a single side effect, their overall risk perception did not differ significantly between the numeric and graph formats ($t(646) = 1.30$, $p = .20$). However, when they read about the combined side effects, those in the numeric condition provided lower risk estimates than both those in the graph condition ($t(646) = -2.81$, $p < .01$).

Using pictographs and a different scenario, this study confirmed the efficacy of presenting risk information in an appropriate graphical format. It appears that visually highlighting the additive nature of multiple risks reduces consumers' tendency to engage in categorical averaging and eliminates the subtractive judgment phenomenon documented across multiple studies.

One important limitation of studies 5A and 5B is that the graphs used to depict multiple risks accurately portray only mutually exclusive risks. However, in reality, multiple risks are not always mutually exclusive. When they are not, accurate graphs would need to account for their overlap. To address this limitation, we conducted an additional study (study 5C, web appendix 10) in which we distinguished between overlapping and mutually exclusive risks. We found that, in both cases, accurate graphs reduce misperceptions of multiple risks.

GENERAL DISCUSSION

Nine studies provide converging evidence of consumers' miscalibrated perceptions of multiple risks. Studies 1A, 1B, and 1C showed that consumers tend to judge combined items with high- and low-risk levels to be less risky overall than the high-risk item alone, and that this difference in perceived risk affects their behavior. Study 2 shed some light on the role of categorical averaging as the process underlying subtractive judgments in the perception of multiple risks. It showed that when consumers are presented with numerical probabilities, they do not in fact attempt to integrate the numbers per se. Instead, they average across the vaguer risk categories that the numbers represent.

Study 3 tested the idea that categorical averaging should be less likely when one side effect is perceived to be too risky. It found that when the probability of one side effect was exceedingly high, adding an SSE neither increased nor decreased consumers' overall risk perceptions, regardless of the probability of the SSE.

Study 4 examined the impact of enticing consumers to rely on verbatim numerical probabilities. It found that when explicitly asked about probabilities as opposed to overall risk, consumers correctly report higher probabilities for the option with two side effects than the option with a single side effect.

The next two studies examined the possible interventions through presentation mode. Study 5A found that when multiple risk information is presented using a bar graph showing the low risk stacked on top of the high risk, consumers do not perceive the combined risks to be lower than the high risk alone. However, presenting the high and low risks as side-by-side bars does not yield the same benefit, as consumers still engage in categorical averaging and report subtractive perceptions of multiple risks. Finally, study 5B investigated the impact of presenting risk information using a different graphical format. It found that pictographs are highly effective at visualizing the additive property of multiple risks, leading to less categorical averaging and eliminating the bias in the perception of multiple risks.

Theoretical Contributions

Our findings contribute to the risk perception and categorical thinking literatures. Extant research on risk perception has focused mainly on the perception of single risks, and when it addressed multiple risks, it tended to emphasize the role of severity. For example, Sivanathan and Kakkar (2017) showed that when a drug lists both minor and severe side effects (vs. only the severe side effects), consumers perceive its overall severity to be lower. Much less is known about how consumers combine probabilities when forming overall risk impressions.

Normatively, an option involving multiple risks should be judged as riskier than an option involving only one of these risks. For instance, a medication that comes with one side effect that occurs in 30% of the people who take it and an SSE that affects 1% of the people who take it is objectively riskier than a medication that comes with only the FSE. However, we show that consumers often perceive the medication with only one side effect to be riskier than the medication with two side effects.

Furthermore, our research contributes to the risk perception literature by identifying categorical averaging as a key mechanism through which consumers form their perception of multiple risks. When combining items with different risk levels, consumers tend to average across the categorical judgments of high risk and low risk, which leads to an overall risk assessment for the combined items that is lower than the perceived risk associated with the high-risk item alone. This discovery can enhance our understanding of why consumers judge an option with multiple potential adverse effects to be less risky than an option with only one of these potential adverse effects.

Finally, our findings extend the literature on information integration by offering novel insights into the averaging process and its boundary conditions. We show that consumers do not average the numerical probabilities they encounter, but the risk categories that these probabilities represent. Our analysis implies that two conditions are necessary for subtractive risk judgments to happen: (1) reliance on categorical representations of risk instead of actual probabilities and (2) use of the averaging rule when combining multiple risks to form overall impressions. This led to the identification of a set of theoretically informed moderators. We reasoned that factors that contribute to the violation of either condition would help eliminate subtractive judgments. Such factors include a side effect that is too risky (study 3), a judgment task that requires the use of verbatim probability information (study 4), and visualization methods that encourage the use additive integration rules (studies 5A and 5B).

Practical Implications, Limitations, and Future Research Directions

This research also offers several practical implications on how to improve consumers' understanding of the risks involved in their health and medical care decisions. Without a good understanding of these risks, consumer engagement is unlikely to achieve the intended benefits of improved quality of care and cost control (IOM 2010). Our findings indicate that presenting multiple risks in nonadditive formats (e.g., numerical format in studies 1–5, side-by-side bars in study 5A) leads to miscalibrated perceptions of these risks; yet, changing the presentation mode to formats that visualize the additive property of multiple risks (e.g., stacked bars in study 5A, pictographs in study 5B) helps correct the misperception. Armed with this information, policymakers can require marketers to present multiple risks in an additive format so that consumers can have a clear understanding of these risks and make informed decisions. For example, Internet-connected medical devices, such as wearables and home health monitoring devices, are highly vulnerable to cyberattacks, which could negatively impact device functionality, data reported to healthcare providers, and emergency notifications. Marketers of these products could be required to disclose those risks using a presentation format that highlights their additive property.

Another approach to reducing the misperception of multiple risks is to educate consumers about the perils of categorical averaging. People are less likely to rely on a cognitive strategy after learning that it leads to erroneous judgments. We tested this intervention in an additional study (study 6, web appendix 11) and found that, compared to a control group, consumers who were made aware of the averaging bias were less likely to commit the averaging error of judging the combination of one high risk and one low risk as overall less risky than just the one high risk.

The current investigation is not without its limitations. First, we were able to investigate the effect of combined risks on risk misperception in limited product categories (e.g., medications, dietary supplements, eye massagers). We cannot say whether our findings would hold for other

categories. For example, processed foods often contain saturated fats, added sugar, and salt and are associated with increased risks of cardiovascular disease, coronary heart disease, and cerebrovascular disorders. Would subtractive risk judgments occur in the context of such multiple risks, and could they be corrected with graphical presentations? Second, participants in our studies were asked to choose for themselves. Previous literature suggests that consumers may process risk information differently when they choose for others (Yang, Saini, and Freling 2015). Future research can examine how choosing for others may affect their evaluation of multiple risks.

DATA COLLECTION INFORMATION

Mehdi Mourali collected the data for studies 1A, 1B, 2, 4, 5A, 5B, and 5C. Zhiyong Yang collected the data for the pilot study and for studies 1C, 3, 3B, and 6. The authors jointly analyzed the data. Data for the Pilot Study were collected on Prolific Academic in May 2022. Data for study 1A were collected on Prolific Academic in February 2020. Data for study 1B were collected on Prolific Academic in January 2021. Data for study 1C were collected on Amazon MTurk in February 2021. Data for study 2 were collected on Prolific Academic in July 2021. Data for study 3 were collected on Prolific Academic in August 2021. Data for study 3B were collected on Amazon MTurk in February 2021. Data for study 4 were collected on Prolific Academic in November 2021. Data for study 5A were collected on Prolific Academic in April 2020. Data for study 5B were collected on Prolific Academic in June 2022. Data for study 5C were collected on Prolific Academic in March 2021. Data for study 6 were collected on Amazon MTurk in March 2021. The data and analysis scripts are available on the Open Science Framework: https://osf.io/bjmhv/?view_only=7d9f8b7cf8e4498c81bac9b24d5b00fa

REFERENCES

Alba, Joseph W. and Howard Marmorstein (1987), "The Effects of Frequency Knowledge on Consumer Decision Making," *Journal of Consumer Research*, 14 (1), 14–25.

Anderson, Norman H. (1965), "Averaging versus Adding as a Stimulus-Combination Rule in Impression Formation," *Journal of Experimental Psychology*, 70 (4), 394–400.

——— (1981), *Foundations of Information Integration Theory*, New York: Academic Press.

Anderson, Norman H. and Gwendolyn R. Alexander (1971), "Choice Test of the Averaging Hypothesis for Information Integration," *Cognitive Psychology*, 2 (3), 313–24.

Ancker, Jessica S., Yalini Senathirajah, Rita Kukafka, and Justin B. Starren (2006), "Design Features of Graphs in Health Risk Communication: A Systematic Review," *Journal of the American Medical Informatics Association*, 13 (6), 608–18.

Aydin, Mehmet A. and Huseyin B. Gokcen (2019), "Factors Considered by Patients in Provider Selection for Elective Specialty Surgery: A Questionnaire-Based Survey," *International Journal of Surgery (London, England)*, 69, 43–8.

Brough, Aaron R. and Alexander Chernev (2012), "When Opposites Detract: Categorical Reasoning and Subtractive Valuations of Product Combinations," *Journal of Consumer Research*, 39 (2), 399–414.

Chernev, Alexander and David Gal (2010), "Categorization Effects in Value Judgments: Averaging Bias in Evaluating Combinations of Vices and Virtues," *Journal of Marketing Research*, 47 (4), 738–47.

Cokely, Edward T., Mirta Galesic, Eric Schulz, Rocio Garcia-Retamero, and Saima Ghazal (2012), "Measuring Risk Literacy: The Berlin Numeracy Test," *Judgment and Decision Making*, 7 (1), 23.

Corbin, Jonathan C., Valerie F. Reyna, Rebecca B. Weldon, and Charles J. Brainerd (2015), "How Reasoning, Judgment, and Decision Making Are Colored by Gist-Based Intuition: A Fuzzy-Trace Theory Approach," *Journal of Applied Research in Memory and Cognition*, 4 (4), 344–55.

Dawes, Robyn M. (1971), "A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making," *American Psychologist*, 26 (2), 180–8.

Fagerlin, Angela, Brian J. Zikmund-Fisher, and Peter A. Ubel (2005), "Cure Me Even If It Kills Me: Preferences for Invasive Cancer Treatment," *Medical Decision Making*, 25 (6), 614–9.

French, D. P., T. M. Marteau, V. Senior, and J. Weinman (2000), "Perceptions of Multiple Risk Factors for Heart Attacks," *Psychological Reports*, 87 (2), 681–7.

Gigerenzer, Gerd and Daniel Goldstein (1996), "Reasoning the Fast and Frugal Way: Models of Bounded Rationality," *Psychological Review*, 103 (4), 650–69.

——— (2011), "The Recognition Heuristic: A Decade of Research," *Judgment and Decision Making*, 6 (1), 100–21.

Gigerenzer, Gerd and Peter Todd; the ABC Research Group (1999), *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.

Gutman, Jonathan (1982), "A Means-End Chain Model Based on Consumer Categorization Processes," *Journal of Marketing*, 46 (2), 60–72.

Hermand, Daniele, Etienne Mullet, and Beatrice Coutelle (1995), "Perception of the Combined Effect of Smoking and Alcohol on Health," *The Journal of Social Psychology*, 135 (2), 167–74.

Hibbard, Judith H. (2004), "New Roles for Patients and Consumers in Assuring High Quality Care," *Virtual Mentor*, 6 (6), 282–4. https://journalofethics.ama-assn.org/sites/journalofethics.ama-assn.org/files/2018-07/msoc1-0406_0.pdf.

Keller, Carmen and Michael Siegrist (2009), "Effect of Risk Communication Formats on Risk Perception Depending on Numeracy," *Medical Decision Making*, 29 (4), 483–90.

Khan, Uzma and Daniella M. Kupor (2017), "Risk (Mis)Perception: When Greater Risk Reduces Risk Valuation," *Journal of Consumer Research*, 43 (5), 769–86.

Leonhardt, James M. and L. Robin Keller (2018), "Do Pictographs Affect Probability Comprehension and Risk Perception of Multiple-Risk Communications?," *Journal of Consumer Affairs*, 52 (3), 756–69.

Lipkus, Isaac M. (2007), "Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations," *Medical Decision Making*, 27 (5), 696–713.

Lipkus, Isaac M. and J. G. Hollands (1999), "The Visual Communication of Risk," *JNCI Monographs*, 1999 (25), 149–63.

Murphy, Gregory L. and Brian H. Ross (1994), "Predictions from Uncertain Categorizations," *Cognitive Psychology*, 27 (2), 148–93.

Newell, Ben R. and David R. Shanks (2003), "Take the Best or Look at the Rest? Factors Influencing "One-Reason" Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (1), 53–65.

Okan, Yasmina, Rocio Garcia-Retamero, Edward T. Cokely, and Antonio Maldonado (2012), "Individual Differences in Graph Literacy: Overcoming Denominator Neglect in Risk Comprehension," *Journal of Behavioral Decision Making*, 25 (4), 390–401.

Peters, Ellen (2020), *Innumeracy in the Wild: Misunderstanding and Misusing Numbers*, New York: Oxford University Press.

Promberger, Marianne and Jonathan Baron (2006), "Do Patients Trust Computers?," *Journal of Behavioral Decision Making*, 19 (5), 455–68.

R Core Team (2021), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/.

Reyna, Valerie F. (2004), "How People Make Decisions That Involve Risk: A Dual-Processes Approach," *Current Directions in Psychological Science*, 13 (2), 60–6.

——— (2008), "A Theory of Medical Decision Making and Health: Fuzzy Trace Theory," *Medical Decision Making*, 28 (6), 850–65.

Reyna, Valerie F. and Mary B. Adam (2003), "Fuzzy-Trace Theory, Risk Communication, and Product Labeling in Sexually Transmitted Diseases," *Risk Analysis*, 23 (2), 325–42.

Reyna, Valerie F. and Charles J. Brainerd (1991), "Fuzzy Trace Theory and Framing Effects in Choice: Gist Extraction, Truncation, and Conversion," *Journal of Behavioral Decision Making*, 4 (4), 249–62.

——— (1994), "The Origins of Probability Judgment: A Review of Data and Theories," in *Subjective Probability*, ed. G. Wright and P. Ayton, New York: John Wiley & Sons, 239–72.

——— (1995), "Fuzzy-Trace Theory: An Interim Synthesis," *Learning and Individual Differences*, 7 (1), 1–75.

——— (2008), "Numeracy, Ratio Bias, and Denominator Neglect in Judgments of Risk and Probability," *Learning and Individual Differences*, 18 (1), 89–107.

Reyna, Valerie F., Wendy L. Nelson, Paul K. Han, and Nathan F. Dieckmann (2009), "How Numeracy Influences Risk Comprehension and Medical Decision Making," *Psychological Bulletin*, 135 (6), 943–73.

Rozin, Paul, Michele Ashmore, and Maureen Markwith (1996), "Lay American Conceptions of Nutrition: Dose Insensitivity, Categorical Thinking, Contagion, and the Monotonic Mind," *Health Psychology*, 15 (6), 438–47.

Shanteau, James (1975), "Averaging versus Multiplying Combination Rules of Inference in Judgment," *Acta Psychologica*, 39 (1), 83–9.

Sivanathan, Niro and Hemant Kakkar (2017), "The Unintended Consequences of Argument Dilution in Direct-to-Consumer Drug Advertisements," *Nature Human Behaviour*, 1 (11), 797–802.

Schirillo, James A. and Eric R. Stone (2005), "The Greater Ability of Graphical versus Numerical Displays to Increase Risk Avoidance Involves a Common Mechanism," *Risk Analysis*, 25 (3), 555–66.

Schwartz, Lisa M., Woloshin Steven, William C. Black, and Welch Gilbert (1997), "The Role of Numeracy in Understanding the Benefit of Screening Mammography," *Annals of Internal Medicine*, 127 (11), 966–72.

Slovic, Paul (2000), *The Perception of Risk*, London: Earthscan Publications Ltd.

Smith, Edward E. and Douglas L. Medin (1981), *Categories and Concepts*, Cambridge, MA: Harvard University Press.

Stone, Eric R., J. Frank Yates, and A. M. Parker (1997), "Effects of Numerical and Graphical Displays on Professed Risk-Taking Behavior," *Journal of Experimental Psychology: Applied*, 3 (4), 243–56.

Stone, Eric R., Winston R. Sieck, Benita E. Bull, J. Frank Yates, Stephanie C. Parks, and Carolyn J. Rush (2003), "Foreground:Background Salience: Explaining the Effects of Graphical Displays on Risk Avoidance," *Organizational Behavior and Human Decision Processes*, 90 (1), 19–36.

Trevena, Lyndal J., Brian J. Zikmund-Fisher, Adrian Edwards, Wolfgang Gaissmaier, Mirta Galesic, Paul K. J. Han, John King, Margaret L. Lawson, Suzanne K. Linder, Isaac Lipkus, Elissa Ozanne, Ellen Peters, Danielle Timmermans, and Steven Woloshin (2013), "Presenting Quantitative Information about Decision Outcomes: A Risk Communication Primer for Patient Decision Aid Developers," *BMC Medical Informatics and Decision Making*, 13 (S2), 1–15.

Troutman, C. Michael and James Shanteau (1976), "Do Consumers Evaluate Products by Adding or Averaging Attribute Information?," *Journal of Consumer Research*, 3 (2), 101.

Visschers, Vivianne H. M., Ree M. Meertens, Wim W. F. Passchier, and Nanne N. K. de Vries (2009), "Probability Information in Risk Communication: A Review of the Literature," *Risk Analysis*, 29 (2), 267–87.

Waters, Erika A., Neil D. Weinstein, Graham A. Colditz, and Karen M. Emmons (2007), "Reducing Aversion to Side Effects in Preventive Medical Treatment Decisions," *Journal of Experimental Psychology. Applied*, 13 (1), 11–21.

Weller, Joshua A., Nathan F. Dieckmann, Martin Tusler, C. K. Mertz, William J. Burns, and Ellen Peters (2013), "Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach: Rasch-Based Numeracy Scale," *Journal of Behavioral Decision Making*, 26 (2), 198–212.

Wilhelms, Evan A., Liana Fraenkel, and Valerie F. Reyna (2018), "Effects of Probabilities, Adverse Outcomes, and Status Quo on Perceived Riskiness of Medications: Testing Explanatory Hypotheses concerning Gist, Worry, and Numeracy," *Applied Cognitive Psychology*, 32 (6), 714–26.

Windschitl, Paul D., René Martin, and Annette R. Flugstad (2002), "Context and the Interpretation of Likelihood Information: The Role of Intergroup Comparisons on Perceived Vulnerability," *Journal of Personality and Social Psychology*, 82 (5), 742–55.

Yang, Zhiyong, Ritesh Saini, and Traci Freling (2015), "How Anxiety Leads to Suboptimal Decisions under Risky Choice Situations," *Risk Analysis*, 35 (10), 1789–800.

IOM (Institute of Medicine) (2010), *The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary*, Washington, DC: National Academies Press (US).

Zikmund-Fisher, Brian J., Peter A. Ubel, Dylan M. Smith, Holly A. Derry, Jennifer B. McClure, Azadeh Stark, Rosemarie K. Pitsch, and Angela Fagerlin (2008), "Communicating Side Effect Risks in a Tamoxifen Prophylaxis Decision Aid: The Debiasing Influence of Pictographs," *Patient Education and Counseling*, 73 (2), 209–14.