Risk Analysis, Vol. 25, No. 3, 2005 DOI: 10.1111/j.1539-6924.2005.00624.x

The Greater Ability of Graphical Versus Numerical Displays to Increase Risk Avoidance Involves a Common Mechanism

James A. Schirillo[1*] and Eric R. Stone[1*]

By displaying a risk reduction of 50% graphically rather than numerically, Stone, Yates, and Parker significantly increased professed risk-avoidant behavior. The current experiments replicated this effect at various risk ratios. Specifically, participants were willing to spend more money to reduce a risk when the risk information was displayed by asterisks rather than by numbers for risk-reduction ratios ranging from 3% to 97%. Transforming the amount participants were willing to spend to logarithms significantly improved a linear fit to the data, suggesting that participants convert this variable within the decision-making process. Moreover, a log-linear model affords an exceptional fit to both the graphical and numerical data, suggesting that a graphical presentation elicits the same decision-making mechanism as does the numerical display. In addition, the data also suggest that each person removed from harm is weighted more by some additional factor in the graphical compared to the numerical presentations.

KEY WORDS: Decision making; risk communication; risk perception

1. INTRODUCTION

One common goal of risk communication is to modify risk-relevant behavior.[1] This task is especially difficult when attempting to convey low-risk probabilities. For example, when probabilities fall below 10%, people do not make decisions according to the expected value theory.[2] However, Stone, Yates, and Parker[3] have shown that framing low-risk probabilities graphically rather than numerically can significantly reduce professed risk-taking behavior. In particular, Stone *et al.*[3] presented their participants with a scenario whereby a manufacturer was considering marketing a new brand of tires. These tires were identical to the manufacturer's standard product except that they reduced the risk associated with tire blowouts by a given amount. Participants were then given the Standard Tire cost and asked how much they would be willing to pay for the safer, "Improved Tires." The risk information was presented either numerically or graphically (e.g., by asterisks). In the numerical condition, the number of people expected to suffer a serious injury was 30 in the Standard Tires condition and 15 in the Improved Tires condition, out of every 5,000,000 Michigan drivers. The graphical condition displayed the values of 30 and 15 in graphs using asterisks (Fig. 1—top row, left columns). Although participants reported they would pay an additional 34% above the stated price when shown numbers (Fig. 1—top row, right columns), they were willing to pay significantly more (i.e., 48% above the stated price) when shown asterisks.

Stone *et al.*'s[3] results are important for both applied and theoretical reasons. On the applied

[1] Department of Psychology, Wake Forest University, Winston-Salem, NC 27109, USA.

* Address correspondence to James A. Schirillo, Department of Psychology, Wake Forest University, P.O. Box 7778, Reynolda Station, 428 Greene Hall, Winston-Salem, NC 27109, USA; tel: (336) 758-4233; fax: (336) 758-4733; schirija@wfu.edu; or to Eric R. Stone, Department of Psychology, Wake Forest University, P.O. Box 7778, Reynolda Station, 222 Greene Hall, Winston-Salem, NC 27109, USA; tel: (336) 758-5729; fax: (336) 758-4733; estone@wfu.edu.

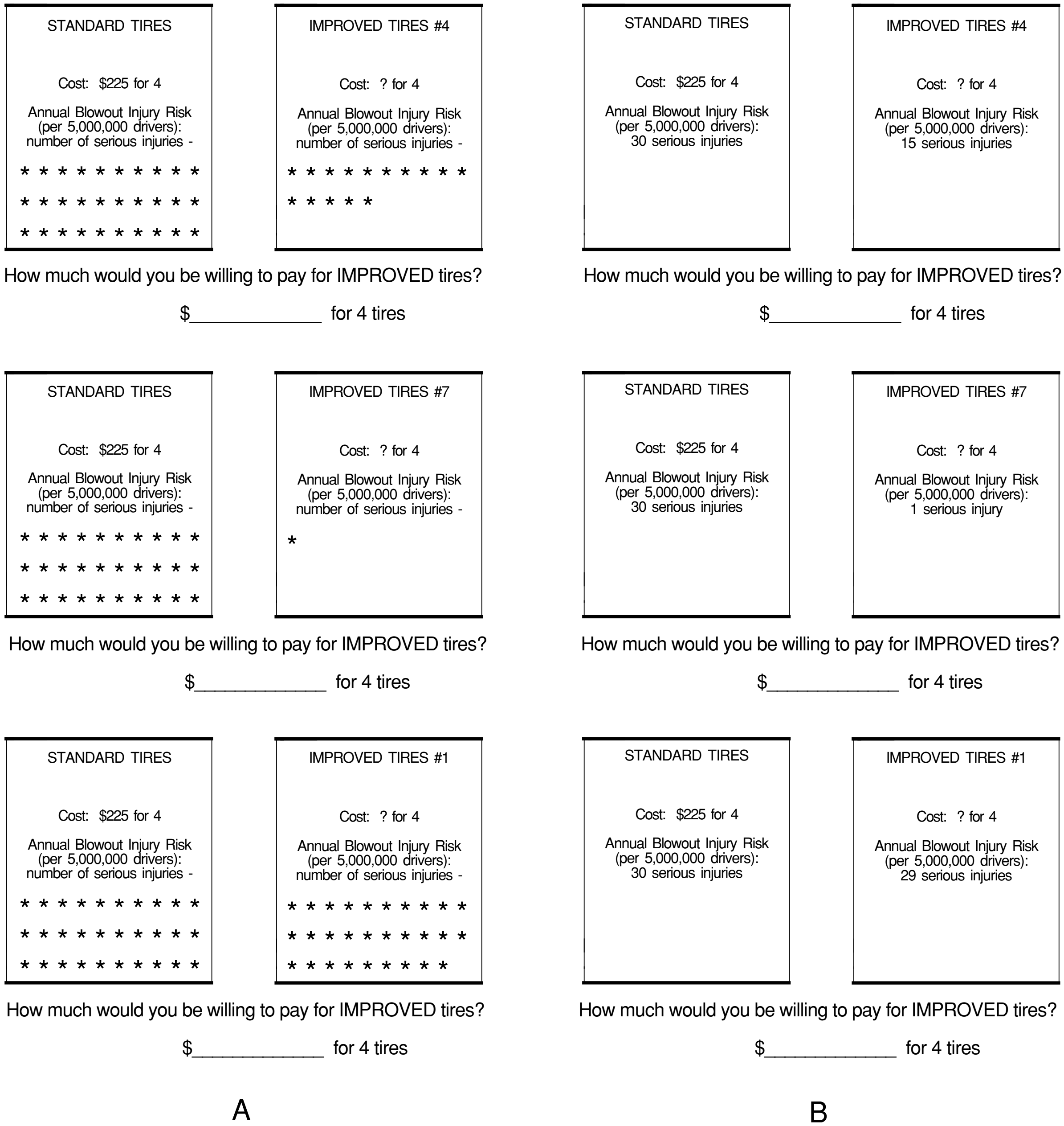

Fig. 1. Box summaries presented to participants in the (A) asterisk format; (B) numbers format (showing only the 15, 1, and 29 injury conditions). Numbers in the boxes (i.e., 4, 7, 1) indicate one Latin-square presentation order.

level, they suggest that risk communicators interested in decreasing risk-taking behavior should seriously consider presenting their risk information graphically (such as by asterisks) rather than numerically. This advice is consistent with that given by other risk communication researchers,[4] but to our knowledge provides its first empirical justification. Theoretically, the differential impact of graphical and numerical information suggests that these two decision-making processes may evoke different mechanisms.

Before accepting these implications, however, it is important to understand what situations Stone *et al.*[3] examined, as so doing might suggest their effect has some boundary conditions. In particular, although the graphical effect appears to hold for a range of different products and associated incidence rates (as long as they are small), this effect has been documented only for risk reductions of 50%. This restriction is potentially problematic, in that one mechanism Stone *et al.*[3] postulated to account for their results posits a pivotal role for graphical depictions of risk reductions of 50% in particular.

Specifically, the explanation provided by Stone *et al.*[3] contained two key components. First, they suggested that risk reductions of 50% are particularly apt to be represented in relative-risk form. As shown by Jarvenpaa,[5] presenting information graphically (as opposed to numerically) makes people particularly apt to attend to the salient aspects of the graphical presentation. As evident in Fig. 1 (top left columns), one particularly salient aspect of Stone *et al.*'s[3] graphical depiction is that the safer product reduced the risk by half. To the extent that graphical depictions do make the 50% risk reduction salient, it is plausible that participants might represent the risk reduction in relative-risk form (i.e., one-half the standard product amount) when presented information graphically. Second, previous research conducted by Stone, Yates, and Parker[6] showed that presenting risk-reduction information using a relative-risk format (i.e., a 50% risk reduction) was more effective in increasing risk avoidance than simply providing incidence-rate information. Drawing on work from fuzzy trace theory[7] and prospect theory,[8] Stone *et al.*[6] hypothesized that when presented with risk information in incidence-rate form, participants represented the risk reduction as "essentially nil" due to the small absolute risk levels involved, but saw the risk reduction as being significant when told that the risk levels had been reduced by half.[9]

Thus, it seems plausible that relative-risk representations are more likely when 50% risk reductions are displayed in graphical versus numerical form, and that this relative-risk representation is a particularly effective means of increasing risk avoidance. To the extent that risk-reduction percentages other than 50% do not make the percentage risk reduction as salient (and thus are less apt to lead to risk reduction being represented in relative-risk form), it is possible that they would not be as effective in increasing risk-avoidant behavior. In particular, using risk reductions whose percentage level is difficult to determine might reduce the graphical effect identified by Stone *et al.*[3] Indeed, although we know of no research that systematically examined different percentage risk reductions, there is evidence that 50% is treated differently than other percentages when people provide probability judgments,[10,11] albeit for reasons different than those hypothesized here. In sum, it is plausible that 50% plays a unique role in how people interpret stated percentage risk reductions, making the impact of graphical formats different at other percentage risk-reduction levels. Put differently, if presented with multiple levels of risk reduction graphically, participants' responses might exhibit a nonlinear response pattern due to a discontinuity at the 50% reduction level. This would produce a nonlinearity in the graphical display results that would be absent in comparable numerical presentations, since the latter format does not emphasize the percentage risk reduction.

Another phenomenon that might result in discontinuous data in graphical displays is called *subitizing*, which proposes that individuals can accurately apprehend only a relatively small number of objects, after which they estimate.[12,13] This process suggests that estimating the value of a small number of objects is more accurate than computing a larger number of objects. In particular, Kaufman *et al.*[13] (their Figs. 6 and 7; here Fig. 2, top and bottom, respectively) show a deviation from a near-perfect correlation between the actual number of dots shown and the reported number of dots above ∼15, such that participants systematically underestimate the number of presented dots at numbers greater than 15.

This process implies that, for multiple levels of risk reduction presented via a graphical display, the slope of a line of each person injured with the improved product from 1 through 15 versus willingness to pay should be relatively constant, but then gradually increase with each additional person injured greater than 15, thus producing a nonlinearity. In contrast, it is reasonable to expect a linear fit in the numbers display condition. In sum, if risk reduction at a number of relative-risk ratios produces deviations from linearity in the graphical compared to the numerical data, different mechanisms may underlie how risk reduction is computed in these two formats.

Conversely, as suggested by an anonymous reviewer, it is plausible that there would be nonlinearity with both the graphical and numerical displays. In particular, prospect theory would predict the greatest changes in perceptions of risk reductions to occur at the extreme margins, so going from 3% to 0% risk or from 100% to 97% would elicit dramatic increases in what participants report being willing to spend. Moreover, this prediction need not presume any increase in the saliency of risk reduction with graphical displays versus numbers at any risk ratio. Thus, if this mechanism is correct, we would expect discontinuities in the graphs for *both* numerical and graphical displays, but with the discontinuities greater at the extreme ends rather than at 50%.

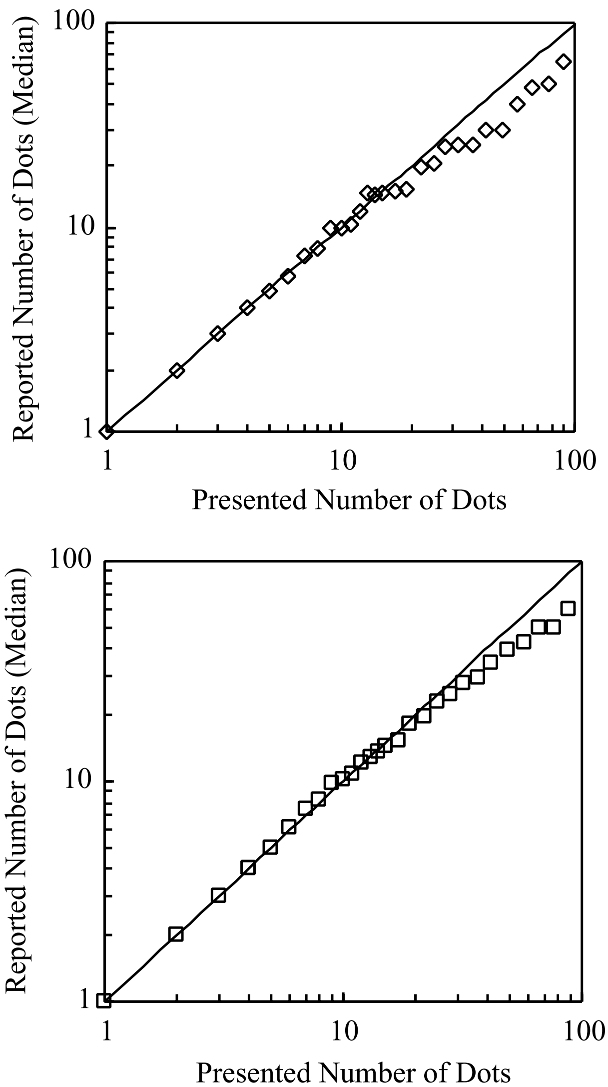

Fig. 2. Median number of dots reported as a function of the number of dots presented. From Kaufman *et al.* (Reference 13) (their Figs. 6 (top—task accuracy with no time constraints) and 7 (bottom—task accuracy in a speeded condition)).

In summary, numerical and graphical formats may either share some common decision-making processes or else activate different decision-making mechanisms. For example, a common process would be evident if people transform one or both variables (i.e., money spent and/or number of serious injuries) to a psychological dimension, regardless of the presentation format. Evidence for different graphical and numerical decision-making mechanisms would be provided if, for example, participants transform the number of injuries in only the graphical condition, where the representation is suspected of highlighting the relative reduction in risk.

Hence, the present studies varied the relative risk ratios used by Stone *et al.*[3] to accomplish two primary objectives. The first was to determine whether or not the graphical effect identified by Stone *et al.*[3] would generalize to risk-reduction ratios other than 50%. From an applied standpoint, ascertaining the generality of the graphical effect is important for determining the effective range of situations to which graphical display formats would apply. Generalizability was also explored by informing participants that the risk of tire blowout was determined for a population of either 50,000 or 5,000,000 drivers. This manipulation is significant in that although the risk-reduction ratios remain identical regardless of the size of the denominator, the denominator does affect the absolute risk reductions.[2] Finding a consistent pattern of results across studies that use denominators varying by a factor of 100 would imply that any documented effects would hold across a range of situations.

Second, testing the generality of the graphical effect may elucidate its underlying mechanism(s). For example, if the graphical effect is considerably stronger at some risk-reduction ratios than at others, as suggested above, different mechanisms might be employed in processing graphical and numerical formats. However, to the extent that the graphical effect holds more generally, the likelihood increases that the two display formats' processing mechanisms are similar. That is, if the equations have the same form for both numbers and graphs across risk ratios, the law of parsimony makes it unnecessary to postulate additional mechanisms for either format. Furthermore, if the asterisks displays increase participants' willingness to pay by a fixed percentage over the numbers displays, there need only be a single mechanism, with asterisks simply more weighted by some additional factor.

2. METHODS

2.1. Participants

Experiments 1 and 2 had 157 and 492 participants, respectively. All were Wake Forest University male and female undergraduates, participating in partial fulfillment of an introductory psychology course requirement.

2 We thank an anonymous reviewer for revealing the value of this important manipulation.

2.2. Materials

The materials were based on those used by Stone *et al.*,[3] which were adapted from Viscusi and Magat.[14] We told participants that there was a certain risk of serious injury due to tire blowouts associated with "Standard Tires," and that the manufacturer was considering marketing a new brand of tires. This brand would be identical in all respects to the former product, except that it would reduce the risk by a given amount. Participants were then told the price of the standard brand of tires was $225 and asked how much they would be willing to pay for the safer "Improved Tires." The relevant information was then summarized and presented to each participant in box form, either numerically or by using asterisks.

In Experiment 1, the "Improved" annual serious injury risk was either 1, 5, 10, 15, 20, 25, or 29 out of every 5,000,000 drivers, compared to 30 seriously injured in the "Standard" condition. In the numbers format, participants were provided with information for both the standard and improved products in terms of the number of people out of 5,000,000 who would be expected to suffer a serious injury (see Fig. 1, right column). The asterisks format was identical except that instead of being given a number, participants were shown asterisks corresponding to the number of individuals who would be injured (see Fig. 1, left column).

The information in Experiment 2 was identical to that of Experiment 1, except that the annual serious injury was out of 50,000 (instead of 5,000,000) drivers.

2.3. Procedure

In both experiments participants were randomly assigned to one of two conditions. One group was presented with the risk information in the numbers format and the other group with the information in the asterisks format. In both conditions, the information was presented in a booklet containing all annual injury risk conditions (i.e., 1, 5, 10, 15, 20, 25, and 29), ordered using a Latin-square design. Participants were not informed at the outset of the values of each of the risk conditions that they would see. They had as much time as they needed to fill out the booklet.

3. RESULTS

3.1. Differences Between Asterisks and Numbers

3.1.1. Within-Subjects Analyses

In Stone *et al.*,[3] participants were willing to pay $76 more than the Standard Tires price of $225 for Improved Tires when given numbers, and $107 more when shown asterisks. In the comparable condition (i.e., 15 injuries) in Experiment 1 participants reported they would be willing to pay $52 more with numbers and $60 more with asterisks, while in Experiment 2 participants reported being willing to pay $67 more with numbers and $83 more with asterisks. Furthermore, Table I indicates that for each of the seven

Table I. Mean Amounts Participants Were Willing to Pay for the Safer Product by Condition

	Condition			
	Experiment 1		Experiment 2	
Nos. of Serious Injuries with the Improved Product	Numbers	Asterisks	Numbers	Asterisks
1	344.59 (71.3)	373.47 (109.5)**	377.50 (158.8)	422.48 (158.3)**
5	317.06 (53.7)	333.86 (80.1)*	341.91 (123.0)	373.60 (123.7)**
10	289.50 (42.4)	312.99 (68.4)**	309.30 (90.6)	343.58 (105.4)**
15	276.60 (39.1)	284.73 (39.6)*	291.52 (79.4)	307.81 (80.3)**
20	258.46 (30.1)	275.60 (51.2)**	265.33 (61.2)	291.70 (66.7)**
25	243.04 (19.3)	253.17 (34.6)**	247.54 (48.5)	251.82 (33.4)ns_2
29	231.28 (22.1)	234.55 (19.6)ns_1	232.27 (29.6)	235.62 (26.2)*
Standard Cost	225.00	225.00	225.00	225.00

**Difference between asterisks and numbers is significant at $p < 0.05$.
*Difference between asterisks and numbers is significant at $p < 0.10$.
ns_1 Experiment 1 difference between asterisks and numbers is nonsignificant ($p = 0.165$).
ns_2 Experiment 2 difference between asterisks and numbers is nonsignificant ($p = 0.13$).
Note: Standard deviations in parentheses (sample sizes ranged from 76 to 80 (Experiment 1) and 245 to 246 (Experiment 2)). As discussed later in the article, we subsequently transformed the dollar amounts by taking logarithms. In Experiment 1 the results of all hypothesis tests were qualitatively similar regardless of whether we used dollar amount or log-dollar amount as the dependent measure. However, in Experiment 2 *all* 7 tests became significant at the 0.05 level when using logs rather than the untransformed amounts.

injury rates that participants evaluated, as a group they *always* stated that they would be willing to pay more for the safer product when presented with asterisks than when presented with numbers.

Thus, although making multiple comparisons to some extent reduced participants' willingness to spend more overall, the graphical effect is a robust phenomenon. The reliability of this effect was tested by conducting repeated-measures ANOVAs, with condition (asterisks vs. numbers) as the between-subjects variable and the amount participants were willing to pay at each of the injury rates as the within-subjects variable. Across these seven injury rates, there was a significant effect of display format in Experiment 1, $F(1, 153) = 4.99$, $p = 0.03$, and in Experiment 2, $F(1, 489) = 10.79$, $p = 0.001$. The effect of the display format was not equivalent for each of the injury rates, however, as indicated by a significant display format by injury rate interaction in Experiment 1, $F(6, 918) = 2.25$, $p = 0.04$, and in Experiment 2, $F(6, 2934) = 8.95$, $p < 0.001$.

We next tested whether the display format effect held for each of the injury rate levels individually. As determined by one-tailed independent t-tests, in Experiment 1, four of these tests were significant at the 0.05 level (1, 10, 20, and 25 serious injuries), and two were marginally significant at the 0.10 level (5 and 15 serious injuries). Only at the value of 29 serious injuries was the difference between numbers and asterisks nonsignificant, $t(154) = 0.98$, $p = 0.165$. In Experiment 2, five of these tests were significant at the 0.05 level (1, 5, 10, 15, and 20 serious injuries), and one was marginally significant at the 0.10 level (29 serious injuries). Only at the value of 25 serious injuries was the difference between numbers and asterisks nonsignificant, $t(490) = 1.14$, $p = 0.13$. (But see Table I note.) Overall, then, it appears that the results found by Stone *et al.*[3] with the 50% risk reduction hold for other levels of risk reduction and, if anything, are slightly stronger at these other levels.

3.1.2. Between-Subjects Analyses

Of concern is whether our results could be due in part to the within-subjects nature of our design. Although the primary independent variable (asterisks vs. numbers) was manipulated between subjects, it is possible that varying the number of people harmed within subjects exacerbated any effect that might be present (e.g., perhaps the asterisks presentation distinguished among the number of people harmed in the different conditions more than the numbers display did). To investigate this potential concern, we analyzed the amounts participants were willing to pay solely for their first response (thereby treating the design as being fully between subjects).

Specifically, we conducted an analysis of variance with condition (asterisks vs. numbers) and number of people harmed on the first trial as the independent variables, and amount participants were willing to pay on the first trial as the dependent variable. The main effect of numbers vs. asterisks was highly significant in both Experiments 1 and 2, $F(1, 143) = 8.82$, $p = 0.003$, $F(1, 478) = 27.29$, $p < 0.001$, respectively. As with the within-subjects analysis, there was also a display format by number of people harmed interaction in Experiment 2, $F(6, 478) = 4.00$, $p = 0.001$. However, this interaction was not significant in Experiment 1, $F(6, 143) = 1.09$, $p = 0.37$.

As with the previous analysis, we next compared the difference between numbers and asterisks for each of the seven different numbers of people harmed. Although the inferential results were generally less significant due to the vastly smaller number of participants per condition, the effect sizes were about the same with the between-subjects comparisons as with the within-subjects comparisons. Table I gives the results aggregated over all placements (i.e., whether it was the participant's first response, second response, etc.). In comparison to those results, the between-subjects asterisks-numbers differences (i.e., considering just the first response) were greater in three of the seven cases in Experiment 1, and in four of the seven cases in Experiment 2.

Finally, we compared the average amounts participants were willing to pay across all numbers of people harmed in the asterisks and numbers conditions for both the within-subjects and between-subjects analyses. Considering all responses (i.e., the within-subjects analysis), participants were willing to pay on average \$280.10 with numbers and \$295.52 with asterisks in Experiment 1, and \$295.05 for numbers and \$318.09 for asterisks in Experiment 2. Considering only the first trial and averaging the seven averages (i.e., mean willingness to pay with 1 person injured, with 5 people injured, etc.), participants were willing to pay on average \$279.22 for numbers and \$300.18 with asterisks in Experiment 1 and \$289.07 for numbers and \$326.70 for asterisks in Experiment 2. Thus, if anything, it appears that the between-subjects effect is slightly *stronger* than the within-subjects effect. At the least, however, our findings are clearly not due to the partial within-subjects nature of our design.

3.2. Curve Fitting

As discussed previously, we were interested in determining what the relationship is between the size of risk reduction and participants' willingness to spend to reduce the risk. In particular, determining if the two equations are qualitatively similar or different provides some preliminary evidence as to the degree of similarity between the mechanisms that underlie how participants treat graphical and numerical display formats.

We used the following analytical approach to determine how best to fit the curves for both numbers and asterisks. First, we examined whether logarithmic transformations of either the dollar amount paid or the number of serious injuries increased the fit to the data. Next, we examined whether including polynomial terms in the equation would give a better fit than a linear fit of the (transformed) data. We examined logarithmic transformations and polynomial terms since these seemed the two most likely ways in which a deviation from a linear fit would improve the fit of the data. Of primary interest was whether any lack of linearity would be the same or different for the two display formats.

3.2.1. Logarithmic Transformations

We first determined whether the relationship between the number of injuries and amount spent is best captured by a linear relationship between these variables or by including a logarithmic transformation of one or both of them. This issue was examined by averaging the responses of the participants at each of the seven numbers of injuries (i.e., determining the means at each of the injury numbers). Similarly, we calculated the logs of the amount spent by taking the log of the responses for each of the individuals, and then averaging those logs.[3] This approach produced a data set consisting of the number of injuries (transformed and untransformed), as well as the average amount spent at each of those levels (transformed and untransformed). Next, we computed correlations between the number of serious injuries (either transformed or not) and the amount spent (either transformed or not) separately for the numbers format and the asterisks format. In other words, for each display format we correlated the amount spent with the number of injuries and with the log of the number of injuries, and correlated the log of the amount spent with both possibilities.

The results of the correlation analysis are given in Table II, which reveal a number of interesting findings. First, even when neither variable was transformed, the correlation between amount spent and number of injuries was quite high, specifically, 0.987 for numbers and 0.985 for asterisks in Experiment 1, and 0.988 for numbers and 0.990 for asterisks in Experiment 2.

Second, in both experiments, transforming the number of serious injuries (i.e., taking lg10 (injuries)) markedly reduced the correlations. Transforming the amount spent to the logarithm of amount spent (i.e., lg10 (dollars)), however, improved the fits considerably. In particular, in Experiment 1, taking the logarithm of amount spent increased the correlation for numbers to 0.995, $t(4) = 4.52$, $p = 0.01$, two-tailed, and for asterisks to 0.993, $t(4) = 2.31$, $p = 0.08$, two-tailed; and in Experiment 2 it increased the correlation for numbers to 0.997, $t(4) = 4.53$, $p = 0.01$, two-tailed, and for asterisks to 0.996, $t(4) = 1.32$, $p = 0.26$, two-tailed.[4] Perhaps most importantly, the results are strikingly similar across both the two display formats and the two experiments. Among other implications to be discussed later, this latter result underscores the reliability of the fitting procedure, as different participants comprised the numbers and asterisks groups as well as took part in the two different experiments.

3.2.2. Polynomial Fits

The next step involved determining whether introducing any polynomial terms would improve the fit to the data. Specifically, we used the transformations that had produced the best fits from the previous section, and examined whether squaring or cubing the number of injuries would improve the fit. That is, we used the log of the amount spent as the criterion variable, and began by conducting a regression equation with only number of injuries in the model. Then, we tested whether or not adding a polynomial term to the equation would increase the fit.

[3] There are other ways of determining the log of amount spent. We took the log of the amount spent for each individual and then averaged those means, since we felt that the logarithmic transformation occurred at the individual level. To check the wisdom of this assumption, we also took the mean of the amount spent and then took the logarithm of that mean, which produced similar, but slightly weaker, results for both display formats. Finally, we calculated the logarithm on the number of people who had serious injuries with Improved Tires, rather than on the number of people saved with Improved Tires (Standard – Improved). The latter approach produced considerably weaker results (i.e., lower *R*s) than did our approach.

[4] The lack of significance for asterisks in Experiment 2 is at least in part attributable to the fact that the untransformed correlation was higher in this condition than in the Experiment 2 numbers condition or in either of the Experiment 1 conditions.

Table II. Correlations Between Serious Injuries Using Improved Tires and Willingness to Pay Using and Not Using Logarithmic Transformations

	Condition							
	Experiment 1				Experiment 2			
	Numbers		Asterisks		Numbers		Asterisks	
	Number of Injuries	lg10 (Injuries)	Number of Injuries	Lg10 (Injuries)	Number of Injuries	lg10 (Injuries)	Number of Injuries	lg10 (Injuries)
Mean of dollars	0.987	0.961	0.985	0.961	0.988	0.959	0.990	0.951
lg10 (dollars)	0.995	0.942	0.993	0.936	0.997	0.935	0.996	0.921

Note: Takes log of willingness to pay for each individual and averages those logs, rather than averaging the untransformed willingness to pay and then taking the log of that mean.

The asterisks condition showed only a negligible improvement by adding either a quadratic or a cubic term in Experiment 1 (both tests of improved model fit produced Fs < 1); however, adding a quadratic term in the numbers condition increased the multiple R from 0.995 to 0.998, which was marginally significant, $F(1, 4) = 5.59$, $p = 0.08$. In Experiment 2, though, none of the increases in fit were significant at the 0.10 level.[5] Thus, in only one of eight cases did adding a polynomial term increase the fit to the data using an alpha level of 0.10. Moreover, this quadratic term was only marginally significant in the numbers format, which was the display for which we had less theoretical basis to predict a nonlinear effect. That is, we thought that any discontinuity that would occur at the 50% risk-reduction level would occur in the asterisks condition. As the linear fit of the number of injuries versus the log of the amount spent was still excellent, even without the quadratic term, this log-linear relationship was used for the rest of the analyses (see Fig. 3A, B).

3.3. Effect of Display Format

As discussed previously, the general form of the equation is comparable for both asterisks and numbers. Nonetheless, the actual equations differ, as suggested by the fact that presenting asterisks generally leads to greater professed risk-avoidant behavior. Specifically, in Experiment 1 the best-fitting curve for asterisks is

$$\log_{10}(\text{amount spent}) = 2.553 - 0.0063 \times (\text{serious injuries with Improved Tires}).$$

And the best fitting curve for numbers is

$$\log_{10}(\text{amount spent}) = 2.525 - 0.0057 \times (\text{serious injuries with ImprovedTires}).$$

In Experiment 2 the best-fitting curve for asterisks is

$$\log_{10}(\text{amount spent}) = 2.603 - 0.0080 \times (\text{serious injuries with ImprovedTires}).$$

And the best fitting curve for numbers is

$$\log_{10}(\text{amount spent}) = 2.553 - 0.0066 \times (\text{serious injuries with ImprovedTires}).$$

Given that the equations are different for numbers and asterisks, it is important to determine what part of the equation produces the greater risk avoidance in the asterisks versus the numbers condition. A larger slope in the asterisks condition would indicate that the impact of each additional serious injury is greater when the risk-reduction information is presented as asterisks rather than as numbers. Presumably, the intercept term would be greater in the asterisks condition than in the numbers condition, due to the substantial difference between the conditions when only one person would be seriously injured with Improved Tires. Of greater relevance, then, is whether the two equations differ when all 30 people would be seriously injured, that is, when the Improved Tires save no additional person from serious injury. A difference here would indicate that the use of asterisks leads to a greater willingness to pay regardless of any actual decrease in risk level.

To determine whether there were differences between the slopes and/or intercepts between the two display formats, we conducted one-tailed tests testing

[5] The only hint of evidence for a polynomial term in Experiment 2 occurred in the numbers condition, where the quadratic term approached significance ($F(1, 4) = 4.24$, $p = 0.11$). This effect is the same effect that was marginally significant in Experiment 1, and the beta was going in the same direction as it was in Experiment 1. Given the weak evidence (at best) regarding it, however, we do not discuss it further.

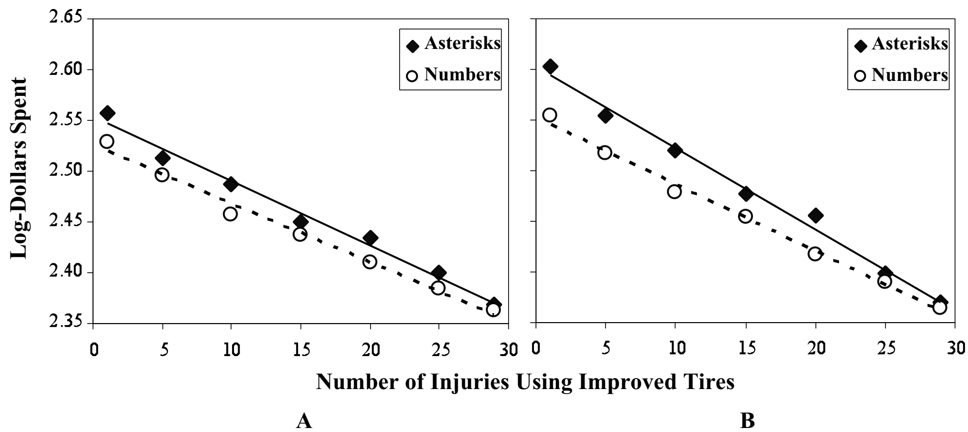

Fig. 3. Log-dollars spent as a function of annual serious injuries associated with using the Improved Tires in the asterisks format (closed diamonds) and in the numbers format (open circles): (A) In Experiment 1 the annual serious injury was out of 5,000,000 drivers, (B) while in Experiment 2 the annual serious injury was out of 50,000 Michigan drivers.

whether the particular regression coefficient was greater in the asterisks display than in the numbers display. In Experiment 1, the slope was greater for asterisks (i.e., −0.0063) than for numbers (i.e., −0.0057), although this difference was only marginally significant, $t(10) = 1.38$, $p = 0.10$, one-tailed. However, in Experiment 2, the slope was significantly greater for asterisks (i.e., −0.0080) than for numbers (i.e., −0.0066), $t(10) = 3.35$, $p = 0.004$, one-tailed. This result suggests that the influence of each additional serious injury was greater when the risk information was represented by asterisks rather than by numbers.

As suggested previously, in Experiment 1 the intercept term was significantly larger in the asterisks condition than in the numbers condition (i.e., 2.553 vs. 2.525, respectively), $t(10) = 3.78$, $p = 0.001$, one-tailed. Similarly, in Experiment 2 the intercept term was significantly larger in the asterisks condition than in the numbers condition (i.e., 2.603 vs. 2.553, respectively), $t(10) = 6.80$, $p < 0.001$, one-tailed. More importantly, the regression equations predicted slightly different log-values at 30 serious injuries in both experiments (i.e., in Experiment 1, 2.364 and 2.354 for asterisks and numbers, respectively; and in Experiment 2, 2.363 and 2.354 for asterisks and numbers, respectively). This difference was only marginally significant in Experiment 1 ($t(10) = 1.45$, $p = 0.09$, one-tailed) and did not quite reach marginal significance in Experiment 2 ($t(10) = 1.15$, $p = 0.14$, one-tailed).[6] Therefore, the greater effectiveness of the asterisks versus the numbers format in inducing risk avoidance appears to be due primarily to a greater influence of each additional serious injury in the asterisks format, though there is a small amount of evidence for a general greater willingness to spend with the asterisks format as well.

The primary cause of the asterisks-numbers difference, then, appears to be the greater effect of each person harmed in the asterisks versus the numerical condition. Of course, simply because there is a statistically significant difference between the two slopes does not necessarily imply that this difference is important in any meaningful way. Thus, it is important to determine the magnitude of the difference between the two slopes. To determine this extent, the slope associated with each person harmed in the asterisks condition was divided by the slope in the numbers condition. In Experiment 1 the ratio of the slopes was 1.11, whereas in Experiment 2 it was 1.21. Thus, each person harmed is given a greater weight of between ∼11% and ∼21% when participants are presented asterisks rather than numbers.

Additionally, the different intercepts at 30 people harmed (i.e., when the safer product does not help anyone) suggests that the use of asterisks leads to a small increase in willingness to pay regardless of any actual decrease in risk level. The extent of this difference was determined by retransforming the log-values into dollar amounts, which showed that in the

6 We tested this difference by computing new regression equations using the number of serious injuries that would be reduced with Improved Tires as the predictor variable. In this case, the intercept term corresponds to no injuries being reduced (and thus to 30 serious injuries). We then tested whether or not the intercept terms were significantly different by using the formulas in Kleinbaum and Kupper.[15]

Experiment 1 numbers condition participants were willing to pay \$225.74 (Standard = \$225) when there was no actual decrease in risk, but \$231.43 in the same situation presented graphically. In Experiment 2 these values were roughly the same as they were in Experiment 1, \$226.04 and \$230.48, respectively.

Finally, we examined the extent to which the greater willingness to pay for the safer product in the asterisks condition was determined by the greater impact of each person harmed in that format versus the general tendency to pay extra regardless of any actual decrease in risk. To evaluate this issue, consider the situation where nobody would be harmed (i.e., the intercept where all 30 people are not harmed with the Improved Tires). Again, retransforming the logs into dollar amounts indicated that in Experiment 1 participants were willing to spend \$335.20 to eliminate the risk when presented with numbers, but \$357.61 when presented with asterisks. In Experiment 2 these values were increased to \$357.02 and \$400.50, respectively. The \$22.41 difference found in Experiment 1 is due both to the general willingness to pay more for asterisks of \$5.69, as well as the greater weight associated with each person saved with asterisks, which produces a \$16.72 difference when all 30 people are removed from harm. In Experiment 2 the numbers-asterisks difference increased to \$43.48, where the general willingness to pay more for asterisks was \$4.44 and the greater weight associated with each person saved with asterisks produced a \$39.04 difference when all 30 people are removed from harm.

Thus it can be seen that, at least in situations where many people are removed from harm, the greater weight associated with each person in the asterisks condition is primarily responsible for the differences between the two display formats. The effect of this greater weight increases as the background information (i.e., the number of people at risk of being harmed) decreases. It seems likely that this greater effect occurs because, with smaller denominators, each person saved is a larger percentage of the population (e.g., 1 of 50,000 is much greater than 1 of 5,000,000). Therefore, it makes sense to pay more to save each person in this situation. To the extent, then, that the asterisks-numbers ratio remains consistent (note that it even increased in Experiment 2), it will produce a greater difference between the two formats.

4. DISCUSSION

Stone *et al.*[3] found that participants reported they would be willing to spend more money for safer tires that reduced serious injuries by 50% if presented that information graphically rather than numerically. The current study demonstrates that this graphical format is more effective than its numerical counterpart in inducing risk-avoidant behavior independent of the relative-risk ratio of serious injuries. As expected, as the number of serious injuries decreased in both the asterisks and numbers conditions (i.e., from 29 to 1), participants increased how much they were willing to spend for the improved product. However, this amount increased at a greater rate in the graphical condition than in the numerical condition. Further, the best-fitting lines in both the asterisks and numbers conditions, although they have different coefficients, are both log-linear (see Fig. 3A, B). The excellent linear fit argues both against any special role played by 50% risk reductions in the graphical condition, as well as against the possibility from prospect theory of a greater change in perceptions of risk reductions occurring at the extreme margins.[8]

In addition, the exceptionally good fits to the current data set with log-linear functions suggest that participants convert the objective number of dollars they would be willing to spend (i.e., numerosity) to its subjective equivalent (i.e., numerousness).[16] Not making a similar psychological translation in the objective number of serious injuries has several important benefits. First, it minimizes the number of mental operations required to determine the extent of risk aversion.[17] That is, taking the log of only one of the two factors cuts the number of possible computations in half. Similarly, not having to take the log of small numbers (i.e., 1–29 lives) reduces the likelihood of misestimating quantities compared to their true values.[18]

More importantly, a parsimonious explanation for the fact that the relationship between the number of serious injuries and the logarithm of willingness to pay was linear for *both* the asterisks and numbers conditions is that a common basic mechanism is used regardless of whether the risk information is conveyed numerically or graphically. This conclusion follows in particular from the fact that the asterisks display is not discontinuous. Having continuity in both data sets further implies that the subjective representation of risk in both conditions is on at least an interval, rather than an ordinal, scale. Although this concept is intuitive in the numerical condition since the external stimuli are presented on an interval scale, this finding strengthens the likelihood that the graphical stimuli also directly map onto a subjective interval scale. Further, this finding suggests that participants were able to subitize the entire range of serious injuries (up to 29 injuries; unlike Kaufman *et al.*'s[13] ∼15 element limit, see Fig. 2).[7] The "neatness" of the graphical arrays may have contributed to our participant's ability to subitize, for example, by facilitating "chunking."[8] Another potential implication of having a common mechanism is that participants' internal representations of the numbers and graphical information might have a straightforward functional relationship. That is, the influence of each person harmed in the two formats could be described by a relatively simple functional relationship, e.g., that the influence is a fixed proportion greater in the graphical than in the numerical condition. It seems doubtful that such a mapping would be as straightforward if the decision-making mechanisms were different for the two displays.

While the current data set follows a log-linear relationship, multiple possible mechanisms could produce such a relationship. For example, it is plausible that participants determine their willingness to spend by examining the number of people saved with the improved product (i.e., number harmed with standard product minus number harmed with improved product). Alternatively, it is possible that they simply examine the number of people being harmed with the improved product, completely ignoring the number harmed with the standard.

Nonetheless, it is worth considering how the present results fit with recent theorizing on potential decision-making mechanisms. Stone *et al.*[19] assumed that people's decisions are influenced by both foreground information (i.e., the number of people affected by some hazard), and background information (i.e., the number of people at risk of being harmed). They postulated that the greater effectiveness of most graphical displays is due to the salience of the foreground to background, whereby graphical formats are more effective because they call attention to the foreground information (which suggests a relatively large risk reduction) and away from the background information (which suggests a relatively small risk reduction, since it shows that the chances of being injured with either product are quite small). To the extent that this explanation is correct, the 11–21% greater "weight" associated with the asterisks presentation may occur because the graphical format emphasizes the foreground information to a greater extent than it does the background information. On the other hand, the mechanism suggested by Stone *et al.*[19] would predict that the ratio would decrease as the denominator decreased, which was not found in the present research. Moreover, Stone *et al.*'s [19] mechanism does not explain why, according to the extrapolated intercept term, the current participants were willing to pay extra when presented with asterisks even when there was no risk reduction. Therefore, it appears that, at a minimum, the explanation provided by Stone *et al.*[19] is incomplete.

4.1. Conclusions and Future Research

Stone and colleagues[3] have already shown that graphical displays can be a powerful aid in conveying risk information that individuals have difficulty imagining or experiencing (i.e., low-probability events). Lipkus and Hollands[20] have specifically requested additional research to identify *how* basic graphical forms affect risk perceptions along such dimensions as magnitude and accuracy of risk. They assert that this type of work would "integrate findings in the disciplines of psychophysics and graphical representations" (p. 160). The current study attempted to do just this. Specifically, graphical presentations, while being more effective than using numbers, do not require an additional transformation into an internal psychological scale prior to decision making. The fact that there is an exceptional log-linear fit to the data for both asterisks and numbers suggests that the two formats utilize a common decision-making mechanism.

Although the approximately fixed percentage increment implies a simple weighting factor in the graphical condition in comparison to the numbers condition, the mechanism of *how* each asterisk gets more weight remains unclear. For example, it is unclear why there is an increase in relative weight (i.e., 11% vs. 21%) when the number of people at risk of being harmed deceases by a factor of 100 between the first and second experiments. In addition, the current findings suggest that future research explore several issues. For example, participants were willing to pay more in the graphical condition than in the numerical condition at zero lives saved. Indeed, the ratio of how much more extra participants were willing to spend beyond the standard price was approximately 8.7:1 in Experiment 1 and approximately 5.3:1 in Experiment 2. These findings suggest that the graphical condition's internal representation may not be on a ratio scale in that it does not share the numerical condition's zero point. Determining why this is the case may elucidate

7 A potential discontinuity remains conceivable, however, in that the numbers quadratic term in Experiment 2 approached significance (see footnote 5).

8 We thank an anonymous reviewer for pointing out this possibility.

why people are generally more affected by graphical than by numerical presentations.

Moreover, although comparable log-linear relations between graphical and numerical conditions implies that the willingness to spend money to avoid low-probability risk is converted into equal psychologically discriminative steps, it has yet to be determined whether this log-transformation occurs prior to or after computing a decision. Determining where in the process the log-transformation occurs would further clarify the structure of the decision-making mechanism.

ACKNOWLEDGMENTS

The authors would like to thank Andrew Parker, Bob Beck, and two anonymous reviewers for their thoughtful comments.

REFERENCES

1. Rohrmann, B. (1992). The evaluation of risk communication effectiveness. *Acta Psychologica*, *81*, 169–192.
2. Fisher, A., McClelland, G. H., & Schulze, W. D. (1989). Communicating risk under title III of SARA: Strategies for explaining very small risks in a community context. *Journal of Air Pollution Control Association*, *39*, 271–276.
3. Stone, E. R., Yates, J. F., & Parker, A. M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, *3*(4), 243–256.
4. Keeney, R. L., & von Winterfeldt, D. (1986). Improving risk communication. *Risk Analysis*, *6*, 417–424.
5. Jarvenpaa, S. L. (1990). Graphic display in decision making—The visual salience effect. *Journal of Behavioral Decision Making*, *3*, 247–262.
6. Stone, E. R., Yates, J. F., & Parker, A. M. (1994). Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavior and Human Decision Processes*, *60*, 387–408.
7. Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, *4*, 249–262.
8. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.
9. Halpern, D. F., Blackman, S., & Goldstein, B. D. (1989). Using statistical risk information to access oral contraceptive safety. *Applied Cognitive Psychology*, *3*, 251–260.
10. Bruine de Bruin, W., Fischhoff, B., Millstein, S. G., & Halpern-Felsher, B. (2000). Verbal and numerical expressions of probability: "It's a fifty-fifty chance." *Organizational Behavior and Human Decision Processes*, *81*, 115–131.
11. Fischhoff, B., & Bruine de Bruin, W. (1999). Fifty-fifty = 50%? *Journal of Behavioral Decision Making*, *12*, 149–163.
12. Taves, E. H. (1941). Two mechanisms for the perception of visual numerousness. *Archives of Psychology*, *37*, 1–47.
13. Kaufman, E. I., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*, 498–525.
14. Viscusi, W. K., & Magat, W. A. (1987). *Learning About Risk*. Cambridge, MA: Harvard University Press.
15. Kleinbaum, D. G., & Kupper, L. L. (1978). *Applied Regression Analysis and Other Multivariable Methods*. North Scituate, MA: Duxbury Press.
16. Stevens, S. S. (1939). On the problem of scales for the measurement of psychological magnitudes. *Journal of Unified Science*, *9*, 94–99.
17. Sparrow, J. A. (1989). Graphic displays in information systems: Some data properties influencing the effectiveness of alternative forms. *Behaviour and Information Technology*, *8*, 43–56.
18. Poulton, E. C. (1985). Geometric illusions in reading graphs. *Perception & Psychophysics*, *37*, 543–548.
19. Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, *90*, 19–36.
20. Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs*, *25*, 149–163.