Comparison-Induced Distortion of Probability

Amber N. Bloomfield (abloomfi@depaul.edu) and Jessica M. Choplin (jchoplin@depaul.edu)
DePaul University Department of Psychology
2219 North Kenmore Avenue
Chicago, IL 60614-3504

Abstract

We investigated the role of Comparison-Induced Distortion (CID) in the representation of risk. Providing comparison risks has been shown to affect risk perception: one's own risk seems higher when higher than the average risk (Klein, 1997) and lower when lower than average, independent of objective risk level. In two studies, we examined the effects of a verbal comparison (e.g. "a higher chance") on recall of probabilities very close or far apart in magnitude. We found that compared probabilities further apart than the difference suggested by comparison language are recalled as closer together (the higher risk recalled as lower, the lower risk recalled as higher), and those closer together than this difference are recalled as further apart (the higher risk recalled as higher, the lower risk as lower). Further, we found that participants' recall for their gender's risk levels flipped as a result of this distortion: the objectively lower probability was recalled as higher than the objectively higher probability. Diagnoses of a person of the same gender followed probability recall. Together, these studies demonstrate that the effect of comparison on risk representation depends on the magnitude of the difference between compared probabilities.

Introduction

Comparison-induced distortion theory (CID theory, Choplin, 2007; Choplin & Hummel, 2002) is a theory of attribute evaluation in which language-based magnitude comparisons (e.g., "He is taller than she is") systematically bias how people evaluate, remember, and estimate attribute values. Rusiecki (1985) found that participants interpreted a comparison phrase, such as "taller than," to imply a particular range of differences in magnitude between compared values (e.g. 2-5 inches' difference for the heights of two women), with substantial consistency across participants in the size of the differences inferred. In CID theory, the magnitude of difference inferred from a comparison phrase is called the "comparison-suggested difference" (CSD). The CSD is the implied difference between values that are compared using a particular word (e.g. "taller") or phrase (e.g. "a completely different ballpark"). This difference may be estimated by asking pre-test participants to provide values that are, for instance, "much more" than a comparison value: the difference between the comparison value and the median value provided by these participants is the CSD (Choplin, 2007). Comparison-suggested differences dictate how a verbal comparison will affect the representations of compared values: if two compared values differ by an amount that is less than that implied by the comparison phrase used to contrast them, they will be remembered as further apart in magnitude than they actually are (Choplin, 2007). Similarly, two compared values that differ by more than the CSD will be remembered as closer together in magnitude than they actually were (Choplin, 2007). This effect has been demonstrated with the recalled size of geometric shapes (Choplin & Hummel, 2002), estimates of personal performance (Bloomfield & Choplin, 2009) and actual behavior: the amount of food people eat is affected by the size of a comparison portion displayed prior to their receiving their own portion (Choplin & Motyka, 2007).

To understand the influence of comparison language in value representation, imagine that a speaker tells their listener "Jane is taller than Marjorie" (see Figure 1). If the actual difference between their heights differs from the CSD for "taller" (~2 inches, Choplin, 2007), the comparison may affect recall of both women's heights (Choplin & Hummel, 2002). For example, if Jane is 5'7" and Marjorie is 5'4", this difference (3") exceeds the CSD, and Jane's height may be distorted (and so recalled) to be shorter than 5'7", closer to Marjorie's 5'4" height, while Marjorie's height may be distorted to be taller, closer to Jane's. Conversely, if Jane is 5'5" and Marjorie is 5'4" (1-inch difference), Jane may be recalled as taller and Marjorie recalled as shorter.

Comparison in risk perception

Risk information, such as having a 1/100,000 chance of developing a disease, may not be "evaluable," in that it may be difficult to interpret how worrisome a risk is or how much the risk requires action from this information alone

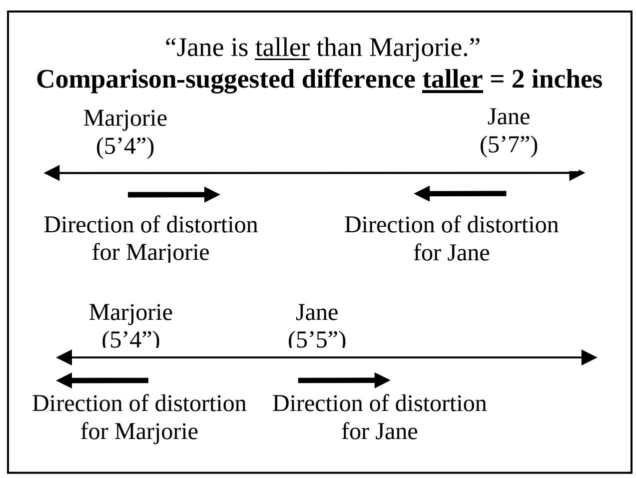

Figure 1. Predicted direction of comparison-induced distortion for greater-than CSD compared values (top) and less-than CSD compared values (bottom).

(Sunstein, 2002). Comparison information is sometimes provided to help people understand their risk (Fagerlin, Zikmund-Fisher & Ubel, 2007; Shoenbach, 1987), but may actually interfere with accurate perception of risk, (Lipkus et al., 2001). For example, relative risk often affects risk perception more than does absolute risk. Klein (1997) found that relative risk had a larger influence on worry judgments than did absolute risk when participants were provided with information about their own risk level (probability of incurring a bad outcome) and information about the average risk level for that same outcome. Participants read that their risk was 20% above or below average, and that their own absolute risk level was 30% or 60% (all factors between-subject). While no significant difference in disturbance or perceived risk ratings emerged between the absolute risk levels (30% vs. 60%), position relative to average affected ratings significantly: participants were more worried about the negative outcome when their risk was above average than when it was below average (Klein, 1997).

Windschitl, Martin & Flugstad (2002) examined the role of relative risk between social groups in risk recall, vulnerability ratings, and hypothetical diagnoses for a member of a particular group (e.g. women vs. men). Participants provided higher ratings and made more diagnoses of a disease which affected the group in question more often than the other group (i.e. had greater *relative* risk for the target group). This pattern of responses continued even when the absolute risk level for the target group was *lower* for the disease with greater relative risk (e.g. when the risk level was 11% for women and 4% for men compared to another affecting 12% of women and 20% of men, participants still gave higher ratings and more diagnoses to the first disease). The authors argued that relative risk impacts the intuitive assessment of risk rather than the estimation of numerical risk (e.g. the probability of acquiring a disease).

Comparison-induced distortion affects recall of compared values across a variety of domains (Choplin, 2007). While the effects of risk comparison have been investigated in a number of studies (Klein, 1997; Fagerlin, et al., 2007; Harris & Smith, 2005; Lipkus et al., 2001; Windschitl et al., 2002), no study has looked at the relationship between magnitude of difference between compared risk levels and the magnitude of difference suggested by the comparison language on risk perception. If comparison-induced distortion occurs for compared risk levels, the introduction of a comparison risk could lead a target risk level to be recalled as higher *or* lower depending on the magnitude of the difference between the compared values.

Experiment 1

Experiment 1 used four disease scenarios adapted from Windschitl, et al. (2002) to examine the effects of difference between compared values on risk recall. For two diseases, men had a higher chance of acquiring the disease, while women had a higher chance for the other two diseases. We measured the comparison-suggested difference for the verbal comparison used to contrast the risk for men and women. The difference between men and women's risk levels was smaller than this CSD for two diseases and larger for the other two. This manipulation of difference magnitude was predicted to affect recalled risk levels in a pattern consistent with comparison-induced distortion: risk levels separated by a large difference recalled as closer together (the larger risk level recalled as smaller and the small level recalled as larger); risk levels separated by a small difference recalled as further apart (the larger risk level recalled as even larger, the smaller risk level recalled as even smaller).

Methods

Materials

Participants read four disease scenarios (one pair affecting the stomach and one pair affecting the skin) via the computer. All scenarios described fictitious diseases affecting men and women at different levels. For instance:

> *There is a disease called Stomach churn discomfort (SCD). This disease affects a larger percentage of women compared to men. This disease affects about 23.8% of women and 11.2% [22.8%] of men. It involves a sometimes embarrassing allergic reaction that causes the stomach to make a loud churning noise after a moderate amount of sugar has been digested.*

The difference between the risk level for men and women was either large (as when 11.2% was shown for the men's risk level) or small (as when 22.8% was shown). Large differences should be larger than the comparison-suggested difference for "a larger percentage," while small differences should be smaller than this CSD. Pairs of probabilities were counterbalanced for the 4 diseases with the restriction that women always had the same risk level for both diseases in the pair (e.g. 23.8% would be the women's risk level for both stomach diseases). The presented risk levels for men and women are shown in Table 1.

Procedure

Participants read and responded to all materials on the computer. The disease scenarios were randomly presented. After reading each scenario, participants provided a potential causal reason for the gender difference. This measure was intended to encourage deeper processing of the numerical risk information. After reading all scenarios, participants engaged in a distracter task then recalled the risk level for each gender for each disease (e.g. the risk level for men for Stomach Churn Discomfort then the risk level for women for this same disease). After completing all recall items, participants were debriefed and dismissed.

Participants

Participants were 70 undergraduate students at DePaul University who received partial course credit for their participation. The data of two participants were eliminated due to a large number of missing data points.

Results

The difference between the recalled gender's risk levels for a disease was divided by the difference between the presented risk levels to determine whether probabilities were recalled as further apart, closer together or the same as presented. As predicted, participants recalled the probabilities separated by a large difference as closer together (i.e. the difference was recalled as smaller, by one-sample t-test, $t(66) = -10.34$, $p < .001$) and those separated by a small difference as further apart (i.e. the difference was recalled as larger, $t(67) = 5.79$, $p < .001$). This result demonstrates that when the differences between presented risk levels is small (i.e. likely to be smaller than the CSD), the probabilities are recalled as further apart and when the difference between presented probabilities is large (i.e. likely to be larger than the CSD), they are recalled as closer together.

Table 1 displays the average recalled probabilities for each gender for each pair of compared probabilities. Recalled probabilities were also divided by presented probabilities to create a proportion. The magnitude of difference between the genders' risk levels significantly affected whether recalled probabilities for men were larger or smaller than observed probabilities ($F(1,65) = 196.47$, $p < .001$ when men had a lower risk level, $F(1, 59) = 82.86$, $p < .001$ when women had a lower risk level). When the actual probability for men acquiring the disease was higher than that for women, men's probability was recalled as higher than presented when the difference between the genders was small (indicating that the men's probability was distorted *away* from the women's; $t(60) = 4.11$, $p < .001$) and lower when the difference between the genders was large (indicating that the men's probability was distorted *towards* the women's; $t(59) = -9.46$, $p < .001$). Similarly, when the men's actual probability was lower than the women's, that gender's probability was recalled as higher when the difference between genders was large (indicating distortion *towards* the women's risk; $t(65) = 8.07$, $p < .001$) and lower when the difference was small (indicating distortion *away* from the women's probability; $t(67) = -6.87$, $p < .001$).

For women, average recalled probabilities were generally lower than actual probabilities and there was no significant effect of difference size on recall (when women had a higher probability, $F(1, 66) < 1$, *n.s.*; when women had a lower probability, $F(1, 58) < 1$, *n.s.*).

Discussion of Experiment 1

Participants recalled the probabilities for men and women as being further apart when they were presented as close together (a 1% or 1.2% difference) and as closer together when they were presented as further apart (a 12.6% or 19% difference). Analyses revealed that this distortion was due to distortion of the men's probabilities and not the women's: women's probabilities were generally recalled as lower than presented regardless of the size of the difference between the gender's probabilities or which gender had the higher risk. One possible reason for this pattern is that women make up a larger proportion than men of the Psychology subject pool at DePaul, and these participants may have been motivated to recall their own gender's risk level as lower. Unfortunately, gender information was not collected in Experiment 1, so no firm conclusions can be drawn about the role of gender in this pattern. Also, as can be seen from Table 1, women's probabilities were recalled fairly accurately, with 17.6% being recalled as approximately 16%, and 23.8% recalled as approximately 21%. These results indicate that participants may have had greater ease in recalling the women's probabilities because these probabilities were repeated twice across the four diseases (i.e. women had only two unique probabilities across the four diseases, whereas men had four). In Experiment 2, unique risk levels were provided for each gender for each disease.

Difference size	Actual men	Recalled men	Actual women	Recalled women
Large	11.2%	15.49%	23.8%	20.60%
	36.6%	24.16%	17.6%	16.54%
Small	18.8%	23.07%	17.6%	15.97%
	22.8%	16.63%	23.8%	21.15%

Table 1. Presented and recalled probabilities for men and women in Experiment 1.

Experiment 2

Experiment 2 improved on Experiment 1's design. First, all presented probabilities for both genders were unique across the four diseases, decreasing the chance that participants would accurately recall risk levels. Second, men and women received different versions of the experiment: the only information that differed between them was the gender described as having "a higher chance" of acquiring each disease (men read about men having a higher chance, women about women having a higher chance). Third, the comparison-suggested difference for the comparison phrase "a higher chance" was measured with pre-test participants to ensure that the small difference magnitude was smaller than the CSD for this comparison, and the large difference magnitude was larger. Finally, in addition to recalling their gender's risk of acquiring each disease, participants made a forced-choice diagnosis of a member of their own gender who had symptoms consistent with both diseases in a pair (i.e. both stomach or both skin diseases). With this item, we sought to examine the effects of distortion in probability recall on judgments.

Methods

Pre-testing

To determine the comparison-suggested difference for "a higher chance," 119 pre-test participants imagined a risk level for a particular disease that was "a higher chance" than a probability described as the average risk level for the disease. The median response was approximately 7% higher than the average risk level. To examine the effects of comparison-induced distortion on probability recall, "small difference" risk pairs in Experiment 2 are less than 7% different and "large difference" risk pairs are greater than 7% different.

Materials

Participants read four disease scenarios that were simplified versions of diseases used in Experiment 1. Two diseases affected the skin and two affected the stomach. For instance:

There is a disease called Stomach churn discomfort. Women have a higher chance of developing Stomach Churn Discomfort than do men. Women have a 24.7% chance and men have a 19.6% chance of developing Stomach Churn Discomfort.

As in Experiment 1, probabilities for the target gender (women in the example above) and the opposite gender were separated by a large difference (> 7%) or a small difference (< 7%). For the each pair of diseases (stomach-affecting and skin-affecting), one disease was a "small difference" disease, and the other was a "large difference" disease. The "small difference" disease in each pair *always* had an objectively lower chance of affecting the target gender than the "large difference" disease in the pair (see Table 2). Corresponding probabilities for the target and other gender were counterbalanced across the two diseases in the pair (e.g. a particular stomach disease was the "small difference" disease for half the participants and the "large difference" disease for the other half).

After engaging in a distraction task, participants saw the other gender's probability of getting each disease and recalled the target gender's probability (their own gender) for both diseases in the pair. After recalling probabilities for both pairs of diseases, participants made a diagnosis for a hypothetical person of the target gender. For example:

Josie is a female college student. She has been having symptoms that are consistent with both Gluten Bloat Condition and Stomach Churn Discomfort. Her doctor is certain that she has one of these diseases, but not both. Which disease do you think Josie has?

This item pitted the perceived risk levels for the diseases in a pair against one another. In the example above, when *Gluten Bloat Condition* was the "small difference" disease, affecting the target gender at a lower rate, *Stomach Churn Discomfort* was the "large difference" disease, affecting the target gender at a higher rate, and vice-versa. If diagnoses were made based on observed objective rate of occurrence, participants should diagnose the "large difference" disease.

Procedure

Participants read and responded to all measures with pen and paper. The four disease scenarios were presented on separate quarter-sheets, randomly ordered. Following this information was a brief distractor task involving "interesting facts." Participants then saw the other gender's probability of getting each disease within a pair (skin or stomach) and recalled the probabilities for their own gender. After recalling probabilities for both diseases in the pair, participants selected one of the two diseases as the one a hypothetical patient most likely had. Recalled probabilities were collected immediately prior to diagnoses to encourage participants to use probability information to guide their diagnoses. After completing these items for both pairs of diseases, participants were dismissed.

Participants

Participants were 124 individuals who agreed to participate after being approached by the experimenter in public places in the Chicago area. The data of 7 participants were eliminated as outliers.

Results

Table 2 displays the average recalled probability for the target gender (note that probability for the other gender was presented at recall) for each disease.

For each participant, the recalled probability was divided by the actual probability to create a proportion, and proportions were averaged across the two "small difference" and the two "large difference" diseases. Figure 2 displays these proportions for "small" and "large difference" diseases. A repeated measures ANOVA with gender as the between-subjects factor and difference size as the repeated measure revealed a significant effect of difference size ($F(1,115) = 226.78$, $p<.001$), but not of gender ($F(1,115) = .134$, *n.s.*). All analyses are collapsed across gender.

For the "small difference" diseases, proportions tended to be greater than 1, indicated that recalled probabilities were generally larger than actual probabilities ($t(116) = 1.88$, $p < .07$). For the large difference diseases, proportions were significantly *smaller* than 1, indicating that recalled probabilities were smaller than actual probabilities ($t(116) =$

Type	Difference size	Actual (other v. target gender)	Recalled target
Stomach	Small	19.6 v. 24.7%	25.78%
	Large	14.2 v. 28.6%	24.11%
Skin	Small	28.8 v. 32.6%	32.70%
	Large	25.2 v. 38.3%	31.37%

Table 2: Presented probabilities and recalled target probabilities in Experiment 2

16.26, $p < .001$).

The more interesting finding regarding recall can be seen from examining Table 2: participants recalled the objectively lower-probability disease as having a higher probability than the objectively higher-probability disease for both the skin and stomach-affecting pair. Excluding participants who recalled both diseases as having an equal chance of affecting the target gender (4 Ps for the skin diseases, 5 Ps for the stomach diseases), this pattern of was significantly more common ($\chi 2$ (1, N = 113) = 10.79, $p < .01$ for stomach diseases; $\chi 2$ (1, N = 112) = 6.08, $p < .05$ for skin diseases) than was recalling the "large difference" disease as having a higher probability. Because target risk levels for the "small difference" diseases were distorted to be *higher* in representation, these were recalled as higher than the "large difference" diseases whose probabilities were distorted to be *lower* in representation.

Diagnoses for skin diseases followed probability recall: the "small difference" disease was significantly more likely to be diagnosed than the "large difference" disease, $\chi 2$ (1, N = 116) = 5.20, $p < .05$. Participants more often diagnosed the *objectively lower-probability disease*. Participants did not show a trend for diagnosing either the "small difference" or the "large difference" disease for the stomach diseases ($\chi 2$ (1, N = 116) < 1, *n.s.*). Figure 3 shows the counts of diagnosis choices for the disease pairs. Recalling "small difference" disease as having a higher probability than the "large difference" disease was significant for stomach diseases, so the lack of a significant preference in the stomach diagnoses indicates that participants may have been taking other information into account when making this choice (such as whether the probability presented for all women applied fully to college-age women, the group to which the hypothetical patient belonged).

Discussion of Experiment 2

Experiment 2 provides additional evidence that comparison-induced distortion occurs in probability representation and influences how probabilities are recalled. When presented with the other gender's probability for acquiring the disease, participants' recalled probabilities for the target gender (their own gender) were higher when the difference between the target and other gender's probability was small and lower when this difference was large. This pattern can be explained by the representation of the target gender's probability being distorted away from the other gender's probability when the difference between them was smaller than the difference suggested by comparison language and towards the other gender's probability when the difference was larger than this suggested difference. Distortion in probability representation led participants to flip the higher probability and lower probability diseases in recall: the disease with objectively lower probability was recalled as having higher probability than the objectively higher probability disease. This distortion in recall led to a trend for diagnoses of the lower probability disease, indicating that participants often did believe the lower probability disease to be more likely to afflict a member of their gender.

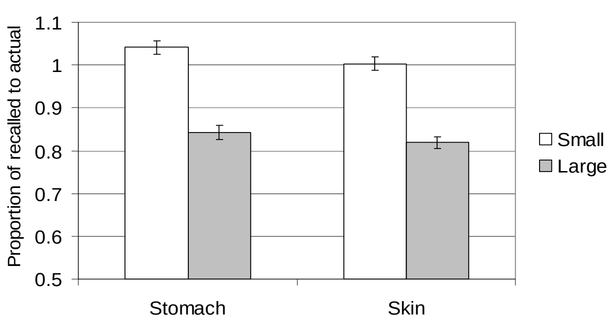

Figure 2: Proportion of recalled to actual probability for small and large difference diseases in Experiment 2

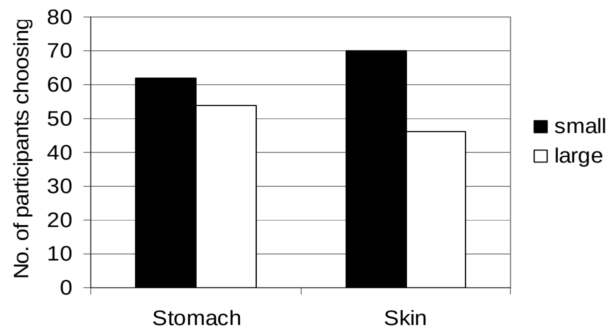

Figure 3. Diagnosis choices ("small" or "large difference" disease) for disease pairs in Experiment 2.

General Discussion

Experiments 1 and 2 demonstrated that comparison-induced distortion (Choplin & Hummel, 2002) affects recall of compared probabilities. When two compared probabilities differ by less than the difference suggested by comparison language (e.g. if the compared probabilities differ by only 3.8%, whereas "a higher chance" implies a difference of about 7%), the representations of these probabilities are distorted to be further apart, the higher probability to be higher and the lower probability to be lower. Conversely, when compared probabilities differ by an amount that exceeds the CSD (e.g. 13.1%), the representations of these probabilities are distorted to be closer together, with the higher probability distorted lower and the lower probability higher. Experiment 2 indicates that distortion in probability representation affects not only recall of these values, but also decisions made based on the recalled probability information. The effect of difference size on diagnosis was also significant for the skin diseases, and this suggests that comparison may affect decisions arising from perceived risk of a disease: participants diagnosed the objective lower probability skin disease that was also the "small difference" disease (and thus distorted by comparison to have a higher probability of affecting the target gender than as presented).

Past research indicates that knowing one has a "higher than average" chance increases concern about the negative outcome in questions (Klein, 1997; Windschitl et al., 2002).

The findings presented here indicate that this is not the only possible outcome of a “higher than average” comparison: if the person’s risk was a large amount higher than the average person’s risk (e.g. if the difference exceeded 7%), her representation of her risk might be distorted down towards the average person’s risk and recalled as lower. Although this distortion might not dissuade the person from being more concerned about this particular risk, since her risk is still “higher than average,” it may play a role in how much time and resources she devotes to one personal risk compared to other risks. For instance, she might recall her risk for developing breast cancer as higher than average but lower than it actually is because average was considerably lower (leading to distortion down *towards* the average risk). She may simultaneously recall her risk of skin cancer as higher than average and higher than it actually is because average was only a small amount lower (leading to distortion up and *away* from the average risk). In deciding where to invest her co-pay costs (i.e. the amount her insurance charges for a doctor’s visit, tests, etc.), she may make an appointment with her dermatologist and forgo a mammogram. In this way, the magnitude of difference between comparison risks can have lead to behavior that opposes what would be dictated by actual risk levels.

The diseases described in Experiments 1 and 2 were fictitious diseases with low severity. In cases where diseases have high severity (i.e. if breast cancer were described) we anticipate the same effect of comparison to appear when a mismatch exists between comparison language and difference between compared values. However, it is possible that the way comparison language is understood (i.e. the comparison-suggested difference for a word or phrase) is influenced by the severity of the outcome in the same manner as interpretation of verbal probability phrases is influenced by severity of an outcome (Harris, Corner & Hahn, 2009; Weber & Hilton, 1990). Verbal probability phrases applied to severe events are sometimes interpreted as referring to a higher probability than when they are applied to less severe events, particularly when the probability information would be instrumental in decisions (Harris et al., 2009). In the case of comparison-suggested differences, a smaller difference in probability may be imagined for “a higher chance” when the comparison is between risk levels for a deadly disease, because even a small increase in risk for this disease would be a cause for concern and warrant the phrase “a higher chance.” We are currently exploring this question and other related questions regarding comparison-induced distortion of risk levels.

Conclusion

The risk level used for a comparison can impact how the comparison affects risk perception. Comparison-induced distortion describes how comparison can influence recall, evaluation, and estimation of compared values. The experiments described here demonstrate that comparison between risk levels is subject to these effects, so the actual magnitude of difference between compared risks must be taken into consideration in order to predict how comparison will impact risk perception.

References

Bloomfield, A. N. & Choplin, J. M. (2009). What it means to be “better:” The role of comparison language in social comparison. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Choplin, J. M. (2007). Toward a comparison-induced distortion theory of judgment and decision making. In J. A. Elsworth (ed.), *Psychology of Decision Making in Education* (pp. 55-86). Hauppauge, NY: Nova Science Publishers, Inc.

Choplin, J. M., & Hummel, J. E. (2002). Magnitude comparisons distort mental representations of magnitude. *JEP: General*, *131*(2), 270-286.

Choplin, J. M. & Motyka, L. (2007). The Interaction of Food-Quantity Differences and Temporal Presentation on the Amount of Food People Consume. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A. (2007). “If I’m better than average, then I’m ok?”: Comparative information influences beliefs about risk and benefits. *Patient Education and Counseling, 69*, 140-144.

Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition, 110*, 51-64.

Harris, P. R. & Smith, V. (2005). When the risks are low: the impact of absolute and comparative information on disturbance and understanding in US and UK samples. *Psychology and Health, 20*(3), 319-330.

Klein, W. (1997). Objective standards are not enough: Affective, self-evaluative, and behavioral responses to social comparison information. *JPSP, 72,* 763-774.

Lipkus, I. M., Biradavolu, M., Fenn, K., Keller, P., Rimer, B. K. (2001). Informing women about their breast cancer risks: Truth and consequences. *Health Communication, 13*(2), 205–226.

Rusiecki, J. (1985). *Adjectives and comparison in English*. New York: Longman.

Schoenbach, V. J. (1987). Appraising the health risk appraisal. *American Journal of Public Health, 77*, 347-352.

Sunstein, C. (2002). Probability neglect: Emotions, worst cases, and law. *The Yale Law Journal, 112*, 61- 107.

Weber, E. U. & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *JEP: Human Perception and Performance, 16*, 781-789.

Windschitl, P., Martin, R., & Flugstad, A. (2002). Context and the interpretation of likelihood information: The role of intergroup comparisons on perceived vulnerability. *JPSP, 82*, 742-755.