

# The generality of the emotion effect on magnitude sensitivity

Min Gong\* and Jonathan Baron

Department of Psychology

University of Pennsylvania

## Abstract

Two studies asked whether reported emotional response interfere with magnitude sensitivity, defined as a subjective evaluation difference between a high magnitude outcome and a low one. Previous research has reported that emotion reduces magnitude sensitivity under separate evaluation in a gain domain (Hsee & Rottenstreich, 2004), a negative effect. We test the generality of this emotion effect in gain and loss domains, and under separate or joint evaluation mode, using a variety of stimuli. We found an opposite, positive, effect in Experiment 1 (willingness to pay to save species or prevent health impairments) but replicated the original negative effect in Experiment 2 (compensation for losses). Further research is needed to disentangle possible causes of these effects and to explore how these findings may be applied to measurement of values for non-market goods.

*Key words:* contingent valuation, emotion, scope test, magnitude effect.

---

\*We thank Fergus Bolger and a reviewer for comments on an earlier version. The work was supported by a grant from the U.S.-Israel Bi-national Science Foundation to J. Baron and I. Ritov. Address: Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104-6241. Email: mingong@psych.upenn.edu or baron@psych.upenn.edu.

# 1 Introduction

In classical economic theory, it is assumed that utility is usually increasing with the magnitude of a good. Although this “the more (good), the merrier” rule seems intuitively self-evident, people do not always follow it. Of interest here, magnitude insensitivity has been reported in various contingent-valuation surveys (Kahneman & Knetch, 1992). Such surveys are used to estimate the value of public goods such as environmental protection, for which no market price exists. Presumably, people ought to be willing to pay more for more protection. Yet, in a typical result, Boyle and his colleagues found “no statistically significant differences in willingness to pay . . . to prevent 2000, 20,000, or 200,000 migratory waterfowl deaths in the Central Flyway” (Boyle et al., 1993). In view of such results, tests of sensitivity to magnitude (scope) are now officially recommended for contingent valuation surveys (Arrow et al., 1993). Yet mere sensitivity to scope is not even sufficient for most purposes, given the fact that policy makers are interested in how much they should spend per bird. Rarely can policy makers tell people the magnitude of the good in each problem to which contingent-valuation results will be applied, so they must infer it from results with other magnitudes (Baron, 1997). The criterion that willingness to pay *per item* is roughly constant as a function of number of items is almost never met. (An exception is Hammitt & Haninger, 2007.)

A question of interest, then, is what makes people more or less sensitive to magnitude of a good when they judge its value. Earlier results suggest that we can think of such judgments as responses to two dimensions: the nature of the good and its quantity. People typically underweigh or ignore the quantity dimension. Instead, they respond in terms of the “prominence” of the good itself (Tversky, Sattath, & Slovic, 1988; Baron & Greene, 1996; Baron, 1997.) More recently, researchers have suggested a role for emotional responses, referring to the dual system theory (Chaiken & Trope, 1999; Kahneman & Frederick, 2002). Specifically, Hsee and Rottenstreich (2004; H&R hereafter) suggested that people used two psychological processes to evaluate a good: valuation by calculation and valuation by feeling. They hypothesized that people were more sensitive to magnitude changes under valuation by calculation than under valuation by feeling. To test this hypothesis, they conducted a series of studies in which they manipulated the emotion evoked by the stimulus and tested whether emotion level interfered with magnitude sensitivity. They found that subjects in the higher emotion group showed lower magnitude sensitivity than those in lower emotion group, which indicated that emotion reduced magnitude sensitivity. In their most informative experiment, they manipulated emotion by showing either 1 or 4 pictures of (endangered) pandas, or, in the control condition, dots.

If the emotion effect found in H&R is general, it can have a significant implication for magnitude sensitivity research. For instance, contingent valuation surveyors usually tend to show participants pictures of goods in question, believing that pictures help conveying information and reducing confusions. However, H&R reported that subjects demonstrated a lower magnitude sensitivity after seeing a picture of an endangered species than if not seeing the picture. The use of pictures may do more harm than good.

But is the emotion effect general and robust? This is the main question we are trying to answer in this paper. There are at least two relevant issues to be investigated. First, in H&R, magnitude sensitivity was tested in separate evaluations (SE hereafter). That is, each subject encountered only one magnitude level. Another widely used valuation mode is joint evaluation (JE hereafter), in which each subject evaluates a series of magnitude levels. Hsee et al. (2005) proposed that “unless the magnitude variable is already highly evaluable under SE, people will be more magnitude sensitive under JE than under SE.” Empirical research also supported this argument (Kahneman et al., 1999; Hsee & Zhang, 2004). So the first question we attempt to answer in the present study is: does the emotion effect observed in separate evaluation apply to joint evaluation? Experiment 1 addresses this question.

Second, all questions in H&R concerned WTP, willingness to pay for a good. Another commonly used measure is WTA, which asks about willingness to accept money for giving up a good. Theoretically, WTP and WTA are two equivalent ways to measure how much people value nonmarketable goods such as environmental problems and health risks. However, WTP measures are typically much lower than WTA measures for the same change in quantity or quality (Knetsch & Sinden, 1984; Baron & Greene, 1996). Since WTA questions are usually in loss domains, and people typically show greater sensitivity to losses than to gains (Kahneman & Tversky, 1979), it is of interest to ask whether the emotion effect on magnitude sensitivity is found in the domain of losses. Experiment 2 is conducted for this purpose.

## 2 Experiment 1: Emotion Effect in Joint and Separate Evaluations

Experiment 1 was conducted to test whether the emotion effect on magnitude sensitivity reported in H&R was a general phenomenon in both joint and separate evaluations for various goods.

### 2.1 Method

The experiment was conducted on the Internet. We sent an invitation by email to 450 members of a larger panel (of 1400) that did other studies. The study was terminated after 85 subjects responded. Two subjects were deleted from the data for answering too fast.<sup>1</sup> Subjects ranged in age from 22 to 70 years with a median of 46 years; 25 subjects were males.

We used eight items, four chosen to evoke positive emotions (endangered species) and four chosen to evoke negative emotions (health risks). Besides the positive/negative category, we also chose individual items so that they would vary in importance to the subjects. The endangered species included: bald eagle, panda, key deer, and Kentucky cave shrimp. Health risks were: losing a toe, paralysis in one leg, paralysis in both legs, and quadriplegia. The health

---

<sup>1</sup>We recorded all times and then formed a time estimate by taking the log of the mean of the fastest half of the page times. These two subjects were outliers on the low side in this distribution.

risks varied in terms of judged severity, according to Gurmankin Levy and Baron (2005). The endangered species also varied in importance, according to a public opinion survey (Coursey, 2001).

Each item had four screens. The first screen contained a verbal description of target item, explaining why it was important to protect the species or prevent the health condition, and it asked subjects to rate the importance of such prevention or protection on a 0–100 scale (“not important at all” to “extremely important”).

Starting at the second screen, subjects were assigned to one of 4 conditions (2 picture presentation X 2 magnitude order) for the item. The picture presentation was either high emotional or low emotional. In the high emotional presentation mode, subjects saw a colorful picture together with a distribution map for the endangered species, and an explicit picture of the health risk. In the low emotional mode, subjects saw only the map for the endangered species, and no picture for the health risk.

In the second screen, below the picture (if present), subjects were asked for their emotional reactions. Emotion was measured by self-report and could be categorized as “immediate emotion” (in the sense of Lowenstein & Lerner, 2003). For the species, the form of the question was positive, e.g., “Please use a number to describe your feeling regarding the bald eagle. The number should be between 0 and 100 where 0 represents ‘have no feeling about it’ and 100 represents ‘very positive’.” For the health problems, the question was negative, e.g., “Please use a number to describe your feeling regarding losing a toe. The number should be between 0 and 100 where 0 represents ‘have no feeling about it’ and 100 represents ‘very negative’.”

Finally, the second screen asked for willingness to pay. An example for the species items is: “Suppose that 140 nesting pairs of bald eagle (2% of the total population) are trapped. The conservation committee is asking 10,000 people (including you) for donations to cover the rescue cost. In your opinion, what is the highest value in the rescue cost X paid (once) by each person that would make you think this is worth rescuing? (Imagine that you might have to pay this for real.)” For the health items, an example is: “What percentage of your annual income are you willing to pay to completely prevent a 35% chance to lose a toe? (For example, if you are willing to pay half of your annual income once, put 50. If twice of your annual income, put 200, and so on.)”

Otherwise, screens 2–4 were all in the same picture presentation condition, and asked only about WTP. They differed in the magnitude of the good. For example, “Suppose that the number of trapped nesting pairs is 350 (5%)?” The magnitude order was counterbalanced so that, for each item, half the subjects saw the highest magnitude first, then the lowest. The other half saw the opposite. (And half of the items that each subject saw went one way, and half the other.) The second magnitude did not appear until the subject had answered the WTP question to the first one. This design was to enable us to test both joint evaluation (within subject) and separate evaluation (between subjects, by looking at the response to the first item in the group of three). The magnitudes used were numbers of animals, in the case of endangered species, and probabilities of getting the condition, in the case of health states. In terms of percent of the population, the magnitudes for the species were: 7% 5% 2% for eagles; 9% 5% 3% for pandas; 8% 5% 2% for

deer; and 5% 3% 1% for shrimp. For the health conditions, the probabilities were: .35 .25 .10 for toe; .45 .25 .15 for leg; .40 .25 .10 for both legs; .45 .25 .10 for quadriplegia. For analysis we consider only the highest and lowest values.

Because the order of magnitudes for each item was counterbalanced across subjects, and because subjects encountered just one magnitude level for the first WTP question and the other questions were on subsequent pages, the difference between the first WTP question when high magnitude was asked and the first question when low magnitude was asked measured magnitude sensitivity in separate evaluation. Magnitude sensitivity in joint evaluation was defined as the log of the WTP ratio (after adding one to all values) for the high and low magnitude within subjects.<sup>2</sup>

## 2.2 Results and Discussion

As designed, subjects regarded the items as having various importance ( $F(3, 328) = 28.51, p < .01$  for endangered species, and  $F(3, 328) = 31.28, p < .01$  for health risks). The picture presentation, however, failed to manipulate the emotion consistently across items.<sup>3</sup> We therefore used rated emotion to test for the emotion effect on magnitude sensitivity. We also computed correlations between emotion and mean WTP (defined as the mean of WTP for high magnitude and low magnitude), and the mean correlation was 0.57 ( $t(80)=16.02, p < .01$  across subjects). The significant positive correlation between mean WTP and emotion indicated that emotion did influence WTP.

We defined magnitude sensitivity as the difference between WTP for high magnitude and low magnitude. Note that in this experiment, WTP was transformed into a logarithmic scale.<sup>4</sup> Subjects showed magnitude sensitivity in both JE and SE. To test the significance of the sensitivities, we defined a subject's magnitude sensitivity in JE as the difference between his mean WTP for all eight high scopes and that for all eight low scopes (which were balanced for order). A subject's magnitude sensitivity in SE was defined as the difference between his mean WTP for all four first high scopes and that for all four first low scopes. Both magnitude sensitivity in JE and in SE were significant ( $t(82)=9.35, p < .01$  in JE,  $t(82)=3.37, p < .01$  in SE). Consistent with previous research (Hsee et al., 1999; Kahneman et al., 1999), subjects showed more magnitude sensitivity in JE than in SE (difference of 0.19,  $t(82)=7.37, p < .01$ ).

To get an idea of the degree of sensitivity, note that the mean ratio (computed from the mean of the logs of each of the eight ratios) of the true scope effects was 3.76; the mean of the JE ratios was 1.40 (computed again from the logs); and the mean of the SE ratios was 1.18. Thus subjects were substantially under-sensitive to magnitude. Part of this was the result of the fact that 54% of the JE responses showed zero sensitivity to magnitude.<sup>5</sup>

So far we have shown that emotion did affect WTP and subjects revealed magnitude sensitivity in both separate and joint evaluations. But was emotion interfering with magnitude sensitivity as suggested by H&R? To answer that

<sup>2</sup>One subject's apparent typo was corrected.

<sup>3</sup>Regressions of emotion rating on importance rating and picture presentation mode (high or low emotion) performed on each of 8 items yielded no significant coefficient before high emotion picture presentation.

<sup>4</sup>To avoid infinite values from  $\log(0)$ , we defined the magnitude sensitivity as the difference between  $\log(WTP+1)$  for the highest magnitude and  $\log(WTP+1)$  for the lowest. We then raised all negative differences to zero, assuming they were mistakes; this did not affect the results.

<sup>5</sup>These included 9% that were originally in the wrong direction.

question, we applied a mixed effects approach (Baayen et al., 2008). This approach, as we used it, is a regression model in which we examine the fixed effect of emotion on magnitude sensitivity, including random effects for individual subjects and individual items, assuming that the effects of subjects and items on magnitude sensitivity are normally distributed, and estimating only the variance of each distribution. Inclusion of these random effects eliminates the problem of non-independence of observations that would result from the fact that each subject does several items and each item is done by several subjects. The lmer function that we used can handle unbalanced designs, such as those that result from missing data (as occurs in Experiment 2). Finally, by removing both random effects simultaneously (as “crossed” random effects) the approach also avoids having to decide whether to find the mean effect for each item and test across items or to find the mean effect for each subject and test across subjects. As we show later (Experiment 2), however, the approach allows us to examine effects of subjects and items separately, within the context of a similar model.

Formally, the model we used to test whether emotion interfered with magnitude sensitivity in joint evaluations can be presented as:

$$MS_{ij} = \beta_0 + \beta_1 Emot_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} \quad (1)$$

where  $MS_{ij}$  is subject  $i$ 's magnitude sensitivity (joint evaluation, defined as the log ratio) on item  $j$ ;  $Emot_{ij}$  is subject  $i$ 's emotion level on item  $j$ ;  $\beta_0$  is the intercept representing the population magnitude sensitivity if there were no emotion effect;  $\beta_1$  is the general emotion effect on magnitude sensitivity, which is what we are interested in;  $\alpha_i$  is the subject random effect (distributed  $N(0, \sigma_\alpha^2)$ ) representing the deviation of each subject from the general magnitude sensitivity;  $\gamma_j$  is the item random effect (distributed  $N(0, \sigma_\gamma^2)$ ) representing the deviation of each item from the general magnitude sensitivity;  $\varepsilon_{ij}$  are model residuals (distributed  $N(0, \sigma^2)$ ); and random variables and model residuals are independent.<sup>6</sup>

Recall that H&R reported a negative interaction between emotion and magnitude sensitivity. Accordingly we would expect to find a negative coefficient for the emotion variable in (1). But instead we found a small, but significant positive one  $\beta_1 = 0.002, p = .0001$ , which indicated that emotion actually helped with magnitude sensitivity in joint evaluations.<sup>7</sup>

We also performed a similar test on data from separate evaluations:

$$WTP_{ij} = \beta_0 + \beta_1 Emot_{ij} + \beta_2 HM_{ij} + \beta_3 HM_{ij} Emot_{ij} + \alpha_i + \gamma_j + \varepsilon_{ij} \quad (2)$$

<sup>6</sup>Note that although we do not have random slopes for emotion level in (1) — thus assuming that the emotion effect did not differ from subject to subject or item to item — we did test on a more complete random effect model in which we included random emotion slopes for subjects and items. But the R package we used (lme4) indicated that the model might have been overparameterized (Baayen et al., 2008). A likelihood ratio test confirmed that and we removed the random slopes.

<sup>7</sup>The p value was not derived from the t value, which was 2.87 (with degrees of freedom undefined). Instead, it was calculated using Markov chain Monte Carlo (MCMC) sampling “from the posterior distribution of the parameters in a Bayesian version of this model” (Baayen et al., 2008). We also tested models with importance rating included, which had no effect, and models in which we included a quadratic component for emotion, which also had no effect.

where  $WTP_{ij}$  was subject  $j$ 's WTP on item  $i$  for the first magnitude;  $HM_{ij}$  is the first magnitude subject  $j$  faced on item  $i$  (1 if the first magnitude was high, 0 otherwise); and other variables and parameters are defined as in (1). We were interested in the interaction coefficient  $\beta_3$ , which informed us whether there was a significant interaction between emotion and magnitude in separate evaluations. We did not find one ( $\beta_3 = 0.002$ ,  $p=0.11$ ), although the direction of the non-significant effect was also positive rather than (as hypothesized) negative.

In summary, in Experiment 1 we found that: 1) emotion had a real effect on the WTP; 2) subjects showed some magnitude sensitivity in both joint evaluation and separate evaluation, but more in joint evaluation; 3) there was a small positive interaction between emotion and magnitude sensitivity in joint evaluation, but not in separate evaluation. To be parallel with the test done by Rottenstreich and Hsee (2004), we also performed tests on 8 individual items separately. One item had an observed p-value of 0.04 and the coefficient was negative for emotion-magnitude interaction in separate evaluation, which might suggest the existence of the negative emotion effect previously reported. But it became non-significant after correction for multiple tests of the same hypothesis.

### 3 Experiment 2. Emotion Effect in the Loss Domain

Experiment 2 was conducted to test whether and how emotion interferes with magnitude sensitivity in the loss domain, where the emotional response was expected to be negative, using different cases and questions.

#### 3.1 Method

Again the experiment was conducted on the Internet. Subjects were from the same subject pool as in Experiment 1, but from a different group. Ninety-three subjects completed the experiment and 5 of them were dropped from finishing the survey too fast. Subjects ranged in age from 21 to 65 years with a median of 46 years; 18 subjects were male. As in Experiment 1, we recorded the time for each response and for each page, and we used the log of the mean of the fastest half of the pages as a measure of each subject's speed.

Following Rottenstreich and Hsee (2001, 2004), we manipulated emotion in Experiment 2 by presenting emotion-rich or emotion-poor instructions. For example, in one of the high emotion settings, subjects saw the following instructions:

In China some people keep chicken as pets. Ding is an eleven-year-old girl who received 5 fluffy chickens from her farmer grandpa as a present. Soon Ding and her chickens became friends. Ding fed the chickens twice a day and the chickens followed Ding everywhere in the yard. Sadly, these 5 chickens died because of an epidemic. Local officials were supposed to have enough medicine to cure the disease, but they ran out and instead were offering to compensate people who lost animals. If you were Ding, how much would

you feel to be a fair compensation for your chickens? Please choose the number that is closest to your answer.

The subject responded by choosing one of 7 buttons, with dollar amounts in equal steps, e.g., “\$20 \$40 ... \$140” for the chicken case. The step sizes were chosen in advance to be suitable for the items. When the subject answered the first WTA question, a second question appeared with the magnitude not used in the first question, e.g., “What if there were 2 chickens instead of 5?” When the subject answered the second magnitude question, the emotion question appeared: “Rate the emotion involved in this case.” The 5-point response scale was labeled “none” and “a great deal” at the two ends.

We used 8 items in Experiment 2: chickens (see above), old books (associated with past people or not), Barbie dolls (formerly owned by landlord vs. daughter), cherries (hand picked or bought), sick dogs to be put to sleep (by a vet who had lost his own dog, or not), house (built by seller or not), museum catalogs (as remembrance of trip, or not), and antique lamps (which reminded owner of mother, or not). Subjects answered questions regarding what compensation or price they would ask for losing or giving up certain amounts (magnitude) of the item, i.e., willingness to accept (WTA). Each item had two levels of magnitude (high or low), e.g., 2 chickens or 5, as well as two levels of emotion. For example, the low-emotion chicken item was:

Ding is an eleven-year-old Chinese girl who lives with her farmer grandpa. To help her grandpa raise 5 chickens on the farm, she fed the chickens twice a day. Unfortunately, all 5 chickens died because of an epidemic. Local officials were supposed to have enough medicine to cure the disease, but they ran out and instead were offering to compensate people who lost animals. If you were Ding, how much would you feel to be a fair compensation for your chickens?

The eight items were presented once each, in a random order chosen for each subject. Four were in the high-emotion condition, four in the low-emotion condition, with the order also randomized for each subject. And half of each set of four were randomly assigned to high-magnitude first or low first. Then the eight items were repeated again, with the emotion condition of each item the opposite of what it had been in the first presentation, but the magnitude order the same; again the order of the 8 items was randomized for each subject. Thus, the two presentations of each item differed in emotion level but not magnitude order; however, because magnitude order was randomly assigned, different subjects saw the same item with a different magnitude order. Results from the two parts did not differ significantly, so the two parts were pooled for analysis. In sum, each subject saw each item at two emotion levels and two magnitude levels.<sup>8</sup>

---

<sup>8</sup>Some items were omitted for some subjects because of a programming error that was corrected in the middle of the study. As we noted earlier, the mixed-effects software we used does not require fully balanced designs.

### 3.2 Results and Discussion

A manipulation check (using the same mixed-effects approach as in Experiment 1) showed that emotion-rich and emotion-poor stories affected the emotion response (difference of 0.86 on the 0–5 emotion scale; *s.e.*=.06). However, we used reported emotion for our main analysis because it was more sensitive.

We found considerable magnitude sensitivity. Recall that subjects answered on a preset scale with numbers. We used the logarithm of the corresponding numbers as the response.<sup>9</sup> As in Experiment 1, we defined joint evaluation (JE) as the log ratio of the responses to the two WTA questions that appeared on the same screen.

We included five additional covariates in models for predicting sensitivity in JE; all were significant correlates of sensitivity in individual analyses and remained significant when included as covariates.<sup>10</sup> First, because of our use of a 7-point scale rather than open-ended responses, responses near the top or bottom of the scale had a limited possible range. Thus, we included the possible range as a covariate in all analyses, using the maximum range as 0 (when the mean of the two responses was 4 on the 7-point scale) and the minimum range (when both responses were at 1 or 7) as  $-6$ . We therefore estimated magnitude effects as the intercept where sensitivity to them was greatest. We also noticed that 8% of the responses used the lowest possible magnitude response for both options, and 10% used the highest. When subjects did this, it is impossible for them to show a JE magnitude effect. Thus, we eliminated these cases (those with a range of  $-6$ ) for all analysis of magnitude effects. Second and third, we found that the effect of magnitude in JE was significantly correlated (positively) with the subject's time, the log of each subject's fastest 8 times (out of 16 pages), and with the log time on each page (truncated at 2 minutes) so we used log mean time per subject and page time (both standardized, so as not to distort the intercept) as covariates in all analyses of the magnitude effect in JE. Fourth, even when we included the subject's mean time, we found that the time on each page was positively correlated with sensitivity (and the subject's time remained significant). Fifth, the ratios of the high and low magnitudes were not the same for the 8 cases, so we included the logs of these ratios (again standardized) as an additional covariate. (We decided not to simply divide by them, because that would assume that subject attended fully to the magnitudes of the numbers, as well as whether they were simply high or low.)

We defined magnitude sensitivity in separate evaluation (SE) in terms of the response to the first WTA question (before the second was presented). Again, we found incomplete magnitude sensitivity, which was greater in JE than SE. Complete sensitivity would yield a mean ratio of 2.09 (across the 8 cases); the mean ratio for JE (based on logs, computed from the model) was 1.67; and the mean ratio for SE was 1.49.<sup>11</sup>

<sup>9</sup>When we simply used the button number, the results were similar.

<sup>10</sup>None of these was significant in Experiment 1, either alone or as covariates, with the exception of time for the individual item, which was, as found here too, positively correlated with emotion.

<sup>11</sup>For the logs, the estimated means were 0.511 (*s.e.*, 0.045) for JE and 0.396 (*s.e.*, .024). It is apparent from the standard errors that these differences are significant.

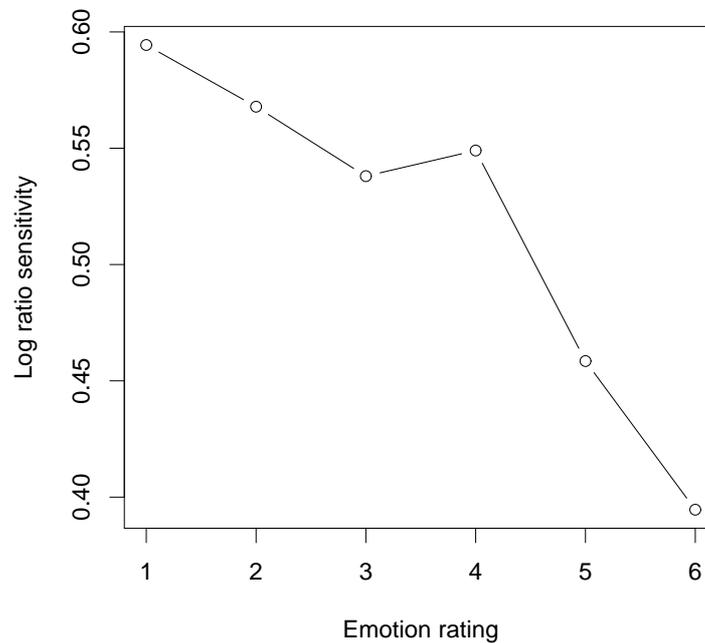


Figure 1: Magnitude sensitivity in joint evaluation (JE) as a function of emotion rating. Estimates are based on the mixed-effects model.

Of greatest interest, the effect of emotion on magnitude sensitivity in JE was clearly negative. The mean effect was  $-.038$ , which means that each step in the 6-point emotion scale reduced the log ratio by this amount, corresponding to a ratio of 0.963 ( $p = .0001$ ).<sup>12</sup> Figure 1 shows the results. Although a quadratic component was significant, when added, it is apparent that this component results from nonlinearity in what is essentially a monotonic effect.

The result was similar when we looked at the effect of emotion-rich vs. emotion-poor items on magnitude sensitivity in JE: the difference between rich and poor was  $-.034$  ( $p = .0238$  by MCMC sampling), corresponding to a ratio of 0.967. Given that the effect of rich-poor on emotion was estimated as  $.864$ , that is, about 86% of one step, and the effect of emotion on sensitivity was  $.963$  per step, the effect of rich-poor is approximately what would be expected if its effect is mediated through emotion.

## 4 Conclusion

Using a mixed effect approach to analyze the interaction between emotion and magnitude under joint and separate evaluation in both gain and loss domain, we found that the negative emotion effect on magnitude sensitivity reported

<sup>12</sup>As in Experiment 1, we used MCMC sampling to test significance. We included cases as a random effect using the 8 cases rather than the 16, because the responses for the high and low emotion versions of each of the 8 cases were highly correlated, and addition of a random effect for nesting of the two versions within each case contributed nothing.

in previous research cannot be generalized to all cases. We found positive effects in Experiment 1, and negative effects in Experiment 2.

The experiments differed in many ways, so we cannot explain these differences. However one possible explanation involves the Yerkes-Dodson Law (Yerkes & Dodson, 1908), which posits an inverted U-shaped relationship between arousal and performance. This empirically relationship has been tested repeatedly in previous research (Anderson, 1990; Revelle & Loftus, 1992). According to Schachter's cognition-arousal theory of emotion, arousal is one of the two factors for an emotional state (Schachter & Singer, 1962), which leads us to suspect the existence of a similar inverted-U relationship between emotion (arousal) and magnitude sensitivity (performance). If we suppose that there is an optimal level of arousal for maximal magnitude sensitivity, then arousal in Experiment 1 would have to be lower than the optimum, and arousal in H&R, and in Experiment 2, would generally be above the optimum. When arousal is below the optimum, increases in emotion will increase magnitude sensitivity, as happened in Experiment 1. When arousal is above the optimum, increases will reduce magnitude sensitivity. The H&R study was conducted in the lab and probably invoked stronger emotion than Experiment 1, which was completed on the Web. The WTA question in Experiment 2 put the problem in a loss domain, which could have elicited stronger emotions toward the objects.

It is possible that the effects of emotion in our experiments are mediated through greater attention. We have no direct measure of attention, but, as noted, we did record the time used to process each page, which ought to serve as a rough measure of attention. Indeed, emotion and time were positively correlated in both experiments; in Experiment 2, the standardized regression coefficient of emotion on time (in a mixed-model analysis) was .11 ( $p = .0001$ ). However, emotion and time had opposite effects on magnitude sensitivity in JE. As noted, emotion reduced sensitivity in Experiment 2, but time increased it (standardized coefficient .10,  $p = .0036$ ). We thus found no support for the possibility that the negative effect of emotion is mediated by attention. Although we did not find a positive correlation between sensitivity and time in Experiment 1, we also did not find a negative one, thus suggesting, again, that time (a proxy for attention) was not the responsible factor for the positive relation between emotion and magnitude sensitivity that we found there.

What is clear is that the emotion effect discovered by H&R is both replicable — albeit in a different context — and not very general. In a situation much like that used in the original study, it reverses.

Our results further muddy the water for contingent valuation surveys. As it is, such surveys already suffer from serious flaws, and other methods for valuing public goods are being used (Baron, 1997). Future studies should look for effects of emotion on these other methods, such as conjoint analysis.

## References

- Anderson, K. J. (1990). Arousal and the inverted-U hypothesis: A critique of Neiss's "Reconceptualizing arousal." *Psychological Bulletin*, *107*, 96–100.
- Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation, *Federal Register*, *58*, 4601–4614.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effect modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Baron, J., & Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*, *2*, 107–125.
- Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin*, *122*, 72–88.
- Boyle, K. J., Desvousges, W. H., Johnson, F. R., Dunford, R. W., & Hudson, S. P. (1994). An investigation of part-whole biases in contingent valuation studies. *Journal of Environmental Economics and Management*, *27*, 64–83.
- Chaiken, S., & Trope, Y., *Dual-Process Theories in Social Psychology*, the Guilford Press, New York, 1999.
- Coursey, D. L. (2001). The revealed demand for a public good: Evidence from endangered & threatened species. In Shogren J. F. & Tschirhart J. (Eds.), *Protecting Endangered Species in the United States: Biological Needs, Political Realities & Economic Choices* (pp. 200–225). Cambridge, MA: Cambridge University Press.
- Gurmankin Levy, A. & Baron, J. (2005). How bad is a 10% chance of losing a toe? Judgments of probabilistic conditions by doctors and laypeople. *Memory & Cognition*, *33*, 1399–1406.
- Hammit, J. K., & Haninger, K. (2007). Willingness to pay for food safety: Sensitivity to duration and severity of illness. *American Journal of Agricultural Economics* *89*, 1170–1175.
- Hanley, N., Shogren, J., & White, B. (1997). *Environmental Economics in Theory and in Practice*. New York: Oxford University Press.
- Heberlein, T. A., Wilson, M. A., Bishop, R. C., & Schaeffer, N. C. (2005). Rethinking the Magnitude Test as a Criterion for Validity in Contingent Valuation. *Journal of Environmental Economics and Management*, *50*, 1–22.
- Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas & muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*, 23–30.
- Hsee, C. K., Rottenstreich, Y., & Xiao, Z. (2005). When more is better: On the relationship between magnitude & subjective value. *Current Trends in Psychological Science*, *14*, 234–237.
- Hsee, C.K. & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology*, *86*, 680–695.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: An Analysis of Decisions Under Risk. *Econometrica*, 47, 313–327.
- Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22, 57–70.
- Kahneman, D., Knetsch J. L., & R. H. Thaler. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives*, 5, 193–206.
- Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk & Uncertainty*, 19, 203–235.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich T., Griffin D., & Kahneman D. (Eds.), *Heuristics & Biases: The Psychology of Intuitive Judgment*, pages 49–81. New York: Cambridge University Press.
- Knetsch, J. & Sinden, J.A., (1984). Willing to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value, *Quarterly Journal of Economics*, 94(3), 507–521.
- Loewenstein, G. & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*, pp. 619–642. New York: Oxford University Press.
- Revelle, W., & Loftus, D. A. (1992). The implications of arousal effects for the study of affect and memory. In S. A. Christianson (Ed.), *Handbook of emotion and memory* (pp. 113-150). Hillsdale, NJ: Erlbaum.
- Rottenstreich, Y. & Hsee, C. K. (2001). Money, kisses, and electric shocks: An affective psychology of risk. *Psychological Science*, 12, 185–190.
- Schachter, S & Singer, J.E. (1962) Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 69, 379–399.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371–384.
- Yerkes, R. M., & Dodson, J. D. (1908.) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology & Psychology*, 18, 459–482.