Abstract

People make judgments and decisions differently depending on whether information that prompts these judgments or decisions is presented in a positive or negative light. This bias is known as the framing effect. Previous research has shown that high numeracy, the ability to understand simple mathematical and probabilistic concepts, makes such framing effects smaller, at least when numbers are involved. Here, we test the possibility that this effect of numeracy occurs because the highly numerate convert information presented in one frame to the other frame. We test the conversions by assessing memory for the frame that was not presented after the judgment has been made. We find that memories of the more numerate (as opposed to the less numerate) are consistent with such conversions, and such memories are associated with smaller framing effects. However these results are moderated by a number of factors: whether the percentage is round or precise, what type of judgment is being made, and how the framing effect is measured. We conclude that the conversions may explain at least a part of the relation between numeracy and framing. Further research is necessary to determine why the roundness of the percentage and the judgment being made affect our results.
Vita

2011..........................................................B.A. Mathematics, Binghamton University
2011..........................................................B.A. Psychology, Binghamton University
2011 to 2012 ..............................................University Graduate Fellow, The Ohio State University
2012 to present ...............................Graduate Teaching Associate, Department of Psychology, The Ohio State University

Fields of Study

Major Field: Psychology
List of Tables

Table 1. Frame X Numeracy interaction performed separately for each item.............. 23

Table 2. Percent of participants answering to the foil of interest correctly (“did not appear”)......................................................................................................................... 24

Table 3. Unstandardized regression coefficients of Model 0. ........................................ 33

Table 4. Crossed random effects models describing effects of numeracy on framing. .... 38

Table 5. Results of Models 2-5........................................................................................... 39

Table 6. Results of Model 3............................................................................................... 40

Table 7. Number of memory errors consistent with conversions .................................. 43

Table 8. Results of Model 7............................................................................................... 44

Table 9. Results of Model 9b................................................................................................ 52

Table 10. Results of Model 10............................................................................................ 53

Table 11. Results of Model 11............................................................................................ 54

Table 12. Means of key variables in Model 11................................................................. 55

Table 13. Four randomly generated order conditions in experiment 1............................ 73

Table 14. Summarized results of Studies 1-3. ................................................................. 75
List of Figures

Figure 1. Mediation Model. ................................................................................................ 4

Figure 2. Numeracy predicts performance on different foils in different ways.............. 18

Figure 3. Means for data corresponding to models Model 3, 4 and 5 (all for the student question) ........................................................................................................................... 41

Figure 4. Correlations and Scatterplots of Individual Difference Variables in Study 1.. 78

Figure 5. Correlations and Scatterplots of Individual Difference Variables in Study 1... 80
Introduction

The framing effect is a well explored phenomenon in judgment and decision making. Three decades of research have shown that consumers, voters and people in general make judgments and decisions differently depending on whether normatively equivalent information that prompts these judgments or decisions is presented in a positive or negative light. In the present study, we investigate attribute framing in particular, which has been deemed the simplest case of framing (Levin, Schneider & Gaeth, 1998). An example of an attribute framing scenario is “John answered 85% of the questions on a test correctly.” A participant is typically asked how well John performed. Judgments following this prompt are compared to judgments in response to the same prompt ending with “15% … incorrectly”. The information in the two prompts is considered normatively equivalent (see McKenzie, 2004, for an alternative viewpoint), but John’s performance is typically rated as superior when participants respond to the former prompt.

Numeracy is the ability to understand arithmetic and probabilistic concepts (see Appendix A for the measure we will use). Peters et al. (2006) found that highly numerate participants (as measured on a subset of this scale), were less susceptible to attribute framing effects and other decision making biases when compared to the less numerate. They suggested that these numeracy differences in the size of attribute framing effects were consistent with the highly numerate being more likely to transform numbers from
one frame to another. Though some research has been done differentiating cognitive mechanisms of the more and less numerate in other decision biases (e.g., Peters & Levin, 2008), no mechanism for how the more numerate avoid being influenced by frame in attribute framing problems has yet been tested. The primary goal of this project is to examine one possible mechanism. A prerequisite to accomplishing this goal is to replicate previous results.

*Hypothesis 1 (replication of Peters et al.):*

The less numerate will show larger attribute framing effects than the more numerate.

The mechanism we propose is a numeric conversion. Specifically, we expect that when presented information in one frame, a participant sometimes converts it to the complementary frame, and that a more numerate participant is more likely to employ such a conversion. For example, when told that “John answered 85% of the questions on the test correctly,” and asked to appraise John’s performance, the more numerate participant may convert this information to “John answered 15% of the questions incorrectly.” Having both frames in mind, the more numerate participant makes a judgment that is less susceptible to the frame in which the information was originally presented.

*Hypothesis 2:*

Conversions of percentages will be associated with smaller framing effects.

This sort of conversion appears likely in light of recent research on differences between more and less numerate people. Cokely and Kelley (2009) conducted a protocol analysis of risky choice. They found that people of high ability, for example the more
numerate, tended to report a number of simple considerations with respect to the probabilities that people of low ability did not. One of these simple considerations was recoding the probability, which is equivalent to the conversions we propose. They did not analyze results for the conversions separately, nor did they examine any framing effects, but they found that making any one of a number of these simple considerations (which include conversions) makes the proportion of risky choices approach what is expected from a normative model. We believe that these conversions are of special importance in attribute framing and therefore look at them separately. We measure the conversions by analyzing memories of the framing scenarios. Specifically, we take false memories of the alternative frame to be evidence that a conversion was made. Therefore, consistent with Cokely and Kelley’s results we expect the following:

**Hypothesis 3:**

More numerate individuals will be more likely to have false memories of the alternative frame, indicating conversions.

Given the Hypotheses 1 through 3, and previous evidence that the more numerate are less biased by frame than the less numerate (Peters et al., 2006), we propose the following mediation hypothesis:

**Hypothesis 4:**

The relation between numeracy and the framing effect will be mediated by false memories of the foils of interest. That is, a part of the reason judgments of the more numerate are less affected by frame is that they make numeric conversions (illustrated in Figure 1).
Note. Higher, compared to lower numeracy was expected to result in smaller framing effects. This relation was expected to be mediated by conversions. Higher numeracy was expected to cause conversions, which would cause smaller framing effects (i.e., more positive judgments when the frame was negative and more negative judgments when the frame was positive).

There are two ways to measure any framing effect: within- and between-subjects. When measuring a between-subject effect, as Peters and colleagues (2006) did, we present each participant with information in only one frame and elicit a judgment; subsequently, we compare judgments of participants who were presented different frames. Alternatively, in a within-subject design, we might present the information in one frame, ask for a judgment, and then after a delay or another task, present the information in the alternative frame, ask for a judgment and subsequently compare the two judgments made by each participant. Some researchers recommend the use of within-subject manipulations for studying framing effects (Levin et al., 2002). These authors suggested
that within-subject designs allow us to correlate framing effects with individual differences. However, within-subject designs are not necessary for this. For example, Peters and colleagues showed that the more numerate have smaller framing effects between-subjects.

More substantially, within-subject manipulations can reveal that some participants are categorically unaffected by frame. For example, without within-subject data on a participant, we cannot tell if a participant considered John’s performance on the test inferior because the information was in a negative frame or because he thought 15% is really a large percentage of questions to answer incorrectly. By contrast, in a within-subject design, we can compare John’s judgment in the positive frame to his judgment in the negative frame and conclude with more certainty that the difference was due to the frame. When examining within-subject data in this way, Levin and colleagues found that some participants do not base their judgments on the frame of the information. However despite their precautions, their findings are also consistent with the possibility that some participants simply remembered their answers in the original frame and in the second part of the experiment responded in the same way, not because they were unaffected by frame, but because they remembered their original answer and perhaps wanted to respond consistently. In other words, they attributed individual differences in their experiment to individual differences in the framing effect, but they may have been caused in part or in whole by individual differences in memory for the original frame.

LeBoeuf and Shafir (2003) argued that this sort of memory effect can be highly problematic. They commented on previous studies that used the within-subject paradigm
to examine another type of framing effect, a risky choice framing effect. These older studies found that framing effects can be made smaller by inducing extended thought. For example, participants asked to provide justifications for their choices showed smaller framing effects (Tetlock, 1992; Miller & Fagley, 1991, Sieck & Yates, 1997) and so did participants who like to engage in thoughtful elaboration, measured by the need for cognition scale (Smith & Levin, 1996). However, LeBoeuf and Shafir find that both of these effects disappeared when the framing effect is measured between-subjects. They go so far as to say “what high NC [need for cognition] respondents are successful at avoiding is inconsistency—not framing per se.” (p. 87, LeBoeuf & Shafir, 2003). It appears useful to differentiate the between- and within-subject framing effects, since according to Shafir and LeBoeuf’s evidence, they may be influenced by different mechanisms. In the rest of the present paper, we specify which type of framing effect we are referring to by differentiating between “within-subject framing” and “between-subject framing.” We will test our hypotheses for each. Since the original findings referred to the between-subject framing effect, we expect our hypotheses to hold for this effect. Since the within-subject framing effects may be driven in part by the realization that the frames are identical and a desire to appear consistent, it appears that differences in within-subject effects should be driven by numeracy and conversions as well.

Within-subject framing effects have been shown to have the highest factor loading in a one-factor model of decision making competence (DMC) that included other decision making abilities such as applying decision rules, consistency in risk perception, resistance to sunk costs, path independence, etc. (Parker & Fischhoff, 2005). Teenagers
with low within-subject framing effects were of higher SES and less likely to employ social services, though the within-subject framing effect had only a directional (non-significant) influence on most other outcomes such as antisocial disorders, externalizing behaviors and delinquency. Bruine de Bruin, Parker and Fischhoff (2007) expanded the knowledge of DMC by creating a more systematic version for use with adults. The results confirmed that a within-subject attribute framing effect has a substantial factor loading on the construct of general decision making competence, but did not examine which outcomes framing effects predict. In sum, the within-subject framing effect may have important real-world correlates and therefore appears to be more than an experimental artifact as Leboeuf and Shafir (2003) had suggested.

Since both the within-subject and between-subject framing effects are important, we examine both. However, given LeBoeuf and Shafir’s (2003) findings that one may be moderated by mechanisms by which the other is not, we do not contend that conclusions about the within-subject framing effect necessarily apply to the between-subject framing effect (or vice-versa), and so in the first study we attempt to address each separately. We further explore between-subject framing effects in Studies 2 and 3 in order to determine if the proposed numeric conversions explain differences in between-subject framing.

**Study 1**

In the first study, we developed an empirical way to test the numeric conversions described above. For the reasons described, we expected that information that is presented in one framed format is sometimes converted to the alternative frame. For example, when the more numerate view the scenario “John answered 85% of the
questions on the test correctly,” and are asked to appraise John’s performance, they may convert this information to “John answered 15% of the questions incorrectly.” Those who make this conversion will thus access both frames, and should make less biased judgments. To measure whether such a conversion has been made, we asked participants to complete a recognition memory task after they made a judgment in response to a scenario. In this task, they viewed some of the words that appeared in the scenario and some that did not. Those who made the conversion should make errors and/or respond more slowly when asked if “15%” (the conversion of the percentage in the original frame) appeared in the preceding scenario. Henceforth, we refer to this as the foil of interest (FOI). Therefore, we should be able to predict the framing effect using subsequent performance on foils of interest, providing an operationalization of Hypothesis 2.

Such a task is usually referred to as the false recognition paradigm (Jacoby & Whitehouse, 1989), in which a conscious or unconscious process inhibits performance on a subsequent memory task. In the original false recognition studies by Jacoby and Whitehouse, participants were shown a set of words to study. After studying the list, they were asked to respond “appeared” or “did not appear” to a set of words, some of which were from the study list (“old”) and some of which were not (“new”). Likelihood of responding “old” was increased by a subliminal prime of that word whether the word was in fact old or new, presumably because its memory was activated by the prime. In our study, we expected that the likelihood of recognition of the foil of interest would be increased by processing the original stimulus, but only if the conversion occurred. This
performance task is useful in the present study because unlike self-reports, it allows access to processes that occur anywhere on the conscious to unconscious continuum. It also has the advantage of being less influenced by demand effects, since the task is framed as a recognition memory task, and unlike in self-report, participants cannot conceal undesirable processes, or invent desirable ones. In other words instead of explicitly reporting mental processes, participants simply try to perform as well as they can on a recognition task.

In the first study, we manipulated the framing effect both within- and between-subjects; we then looked for conversions as a mediating mechanism for the effect of numeracy on each of these framing effects. Participants were asked to make judgments about a diverse range of stimuli in one frame, (e.g., positive: 85% correct), and then after a short break in the alternative frame (e.g., negative: 15% incorrect). Naturally, we can look at differences in judgments to examine the within-subject framing effect. However, randomly varying which frame was presented first between participants also allowed us to measure the between-subject framing effect by examining only answers in the first half of the experiment.

Method

Participants

Seventy-nine state university undergraduates (mean age = 20, 64% female) participated in the study for course credit.
Design

The experiment consisted of 24 framing scenarios for each participant: 4 (scenario category types) x 3 (scenarios in each category) x 2 (frames). Scenarios were presented in the same random order in both frames. About half of the participants were presented the positive frame first, and half the negative. Only the two frames were variables of interest; the 12 scenarios of 4 categories were included to enable us to generalize results beyond one scenario. The four scenario categories were: performance of students based on percentage questions correct (incorrect), healthfulness of snacks based on percentage sugar-free (sugar), nutritional value of milk based on percentage fat-free (fat), and safety of drugs based on percentage of patients experiencing side effects (no side effects). There were three scenarios in each of the four categories (so that each participant responded to 12 scenarios in each frame), with different irrelevant characteristics varied such as names, but also different percentages, hence eliciting different average judgments for different scenarios. One example scenario follows. “In an Introductory Psychology course with 245 students, John took a midterm that covered the history of psychological thought in the first half of the last century. John answered 87% of the questions correctly.” The four framing scenarios that were followed by foils of interest can be found in Appendix B; the rest of the scenarios were structured similarly but involved different details (e.g., different course name, course material, number of students).

Each participant was randomly assigned to one of 4 random order conditions and saw either the positive or negative frame first (see Appendix C for order manipulations).
The order conditions were included to control for order effects, and each participant remained in the same order condition in both frames that the participant viewed.

After each of the 12 framing scenarios in both blocks, participants completed 10 recognition trials. During some of these trials, foils of interest were shown (e.g. “15%” if the original judgment prompt included “85%”). However, in order to prevent participants from “catching on,” only four of the 12 framing scenarios (one in each scenario category) were followed by a foil of interest (one scenario about drugs, milk, snacks and students was followed by a foil of interest). Four other evaluative judgments were followed by other foils, which were randomly generated percentages (e.g., “57%”) used as controls. The eight scenarios followed by foils of interest or other foils were also followed by targets, which were percentages that really did appear in the prompt (e.g., 85%). The same four scenarios that were followed by foils of interest and targets in the first frame a participant encountered were also followed by the foils of interest and targets in the alternative frame. The rest of the 10 recognition trials were filled with distractor items which were words from the scenarios. All hypotheses addressed the foils of interest and some call for comparisons to other numeric foils.

Procedure

After consenting to the experiment, participants read instructions about the upcoming task and completed a practice trial. In the practice trial and in a typical trial in the experiment, they viewed numeric and verbal information about a product or student and made a judgment of quality on a scale from -3 to +3. After each judgment, they proceeded to the recognition phase, in which they were instructed to respond to 10 words
and numbers by indicating “appeared” or “did not appear”. The example in the instructions required participants to distinguish “Egyptian cotton” from “Blended cotton”, explicating that minor details may be tested. The correct response on exactly half of the trials was “appeared” to ensure neither response was biased for or against. Participants made responses on a standard keyboard with index cards over the F and J keys. The F key was labeled “did not appear” and the J key was labeled “appeared.” Each of the recognition trials was preceded by a centered “xXx” and participants were instructed to look at the large X.

After participants completed the 24 quality judgments and the recognition trials corresponding to each in the first frame, they answered vocabulary and demographics questions. These tasks were chosen because they were not expected to influence subsequent judgments, but would provide a time buffer in order to distract participants from remembering the evaluations they had made. In addition, they provided useful tests for confounds (e.g. attentiveness within the experiment, general intelligence) with numeracy when scores were used as covariates in the regression models that follow.

Participants then proceeded to scenarios in the alternative frame, in the same order as in the original frame, followed by recognition trials. The sheer number of framing scenarios and their similar structure were expected to make remembering any particular judgment correctly relatively difficult due to interference, and therefore increase the likelihood of finding inconsistency in judgment.

After both frames were completed, participants completed the full 18 item numeracy measure from which a recent short numeracy scale was developed (Weller et
al., 2012) and a working memory measure. The working memory, vocabulary, and standardized test scores collected in the demographics section were used as proxies for intelligence. Since some of the participants took the ACT and others the SAT, standardized test scores were calculated by converting the observed responses to a normalized metric (the sample mean on both tests were subtracted and the differences were divided by the sample standard deviations). A correlation matrix of these scores is available in Appendix E.

Analysis

Cleaning

Data from three participants were deleted due to extremely low d prime (a measure of discrimination in signal detection theory) in a memory task that was part of the experiment, indicating that they were not paying attention, and from another participant because s/he answered vocabulary questions at chance and reported speaking English for only 1 year. Data from two were lost due to a research assistant’s error, leaving 74 participants for analysis. More data were collected in a second batch toward the end of the semester in order to obtain a sufficient sample size to do a between-subject analysis. End of semester data can be noisy and indeed they were found to be unreliable in this experiment, since there was a significant 3-way interaction between when the data were collected and the numeracy X frame interaction, an important effect in this study. Therefore these data were discarded.
Crossed Random Effects Model

The analyses for this experiment were done using crossed random effects models (Bates, 2010). Similar to other hierarchical mixed effects models, these models generalize multiple regression. A standard mixed effects hierarchical model allows observations for a given participant to correlate, and more accurate estimates for the fixed effects (experimental effects) are obtained. This model, too, accommodates correlations between observations for a particular subject. However, unlike a regular mixed effects model, observations for a given stimulus are also allowed to correlate.1

Dealing With Response Time Outliers

Short outlier response times were not generally a problem. One response was deleted because it was over 200ms faster than the next fastest response for any participant. Ratcliff (1993) suggested that long outliers can result from processes other than those of interest in the study (for example, yawning or other distractions). However, in the present study, the hypothesized conversions were predicted to also result in very long response times. For this reason, instead of deleting data, we used speed (1/RT, as suggested by Ratcliff) in the inferential analysis, and backtransformed the means to RT for the descriptive analysis (reported mean = 1/mean(1/RT)). The distribution of speed was roughly normal.

Results

Numeracy was approximately normally distributed (mean=12.06 and median=12). Since the present sample was relatively highly numerate (college students), we used the
bottom quadrant (numeracy<11) to split between low and high numeracy to present data. However, in the inferential analysis, numeracy was treated as a continuous variable.

The inferential analyses were done using mixed models with random intercepts for participants, and random intercepts for framing scenarios as described above because we had repeated measures for each of these factors. P-values were estimated by bootstrapping as suggested by Baayen, Davidson and Bates (2008), but similar results can be achieved with the additional assumptions of sphericity and compound symmetry in a repeated measure ANOVA. The following analyses concern only the scenarios followed by the foil of interest, unless otherwise stated.

**Within-Subject Framing**

**The Framing Effect.** To test within-subject framing effects, we used the difference between responses in the positive and negative frames for each framing scenario as the dependent variable. In the experiment as a whole, the framing effect was positive on average (M=0.49; t=7.6; p<.01), indicating that participants appraised the same scenario an average of 0.49 points higher on a scale from -3 to +3, in a positive frame than in a negative frame. This difference was predicted by numeracy (t=−2.99, p<.01) such that more numerate participants (M =0.40) were less swayed by frame than the less numerate (M = 0.67) This relation was still significant when controlling for the effects of all intelligence proxies (working memory and vocabulary) (t=−2.6, p<.01). These results confirm previous findings that numeracy ameliorates the framing effect, and extend them to a within-subject paradigm, affirming Hypothesis 1.
Foils of Interest. Recognition of foils of interest after judgments in the second frame should predict the extent of the framing effect, affirming Hypothesis 2. To test this, we fit another crossed random effects model with the framing effect as the dependent variable and recognition of the foil of interest as the independent variable coded 1 when answered correctly (correct rejection) and 0 when answered incorrectly (false alarm). The difference between responses in the positive and negative frames (i.e. the framing effect) was significantly higher when participants answered the foils of interest “did not appear” correctly (M=0.75) than when they answered “appeared” incorrectly (M=-0.15; t= 3.06, p<.01). This finding provides strong evidence that making a conversion to an alternative frame allows a more balanced judgment to be made. Furthermore, the highly numerate erred on about 24% of the foils of interest that followed judgments in the second half of the experiment, while the less numerate erred on only 12% (z=-2.99, p<.01 when controlling for intelligence using proxies), providing evidence supportive of Hypothesis 3.

Testing a Confound. Since our measure of conversions is a memory measure, it may be that having poor memory, not a conversion per se, is responsible for the results reported above. If this is the case, other memory errors should exhibit the same results as above. To test this possibility, we examined responses to another percentage foil that was not the complement of the correct percentage. Responses to this foil did not predict the framing effect (the difference between frames was 0.49 when answered correctly and 0.57 when answered incorrectly; t=-1.05, p>.2), and numeracy did not predict these responses (8% errors in the less numerate and 10% errors in the more numerate; z=-0.49,
p>.2 in the second frame). Thus it is memory for the foils of interest specifically that matter to the framing effects, not memory for numeric information more generally. See Figure 2 for the comparison of results on the foil of interest and the other percentage foil.
Figure 2. Numeracy predicts performance on different foils in different ways. 

*Note.* Mean accuracy rates on recognition of foils of interest and other foils. As predicted, the numerate individuals performed significantly worse on foils of interest (due to conversions) and a little better on other foils (due to their superiority in general number processing). Error bars represent standard errors of the means.
**Response Times.** Slower responses on the foil of interest also predicted smaller framing effects directionally, but this result did not approach significance ($t=-0.98$, $p>.20$). At first glance, it might appear that this is due to the inherent variance in response time measures. However, in unreported analyses, the effect remained nonsignificant even when controlling for a variety of other response time measures, such as response times to other memory trials, or other memory trials of percentages. It appears that participants did not “catch themselves” responding as though the converted information was in the preceding scenario. It is also possible that an analysis of means was not fine grained enough to interpret the response times. Unfortunately, there were not enough data to apply a diffusion model or another model of recognition memory.

**Mediation.** In Hypothesis 4, we predicted that higher (vs lower) numeracy would lead to smaller framing effects at least in part because it leads to conversions (operationalized as errors on the foil of interest) which, in turn, lead to smaller framing effects. When predicting the framing effect, numeracy’s coefficient was $-0.104$ ($t=-2.70$, $p<.01$), but fell to $-0.075$ ($t=-1.96$, $p=.05$) when the dummy variable for answering the foil of interest correctly was included in the regression. Pituch, Whittaker and Stapleton (2005) suggested that mediation tests used for regular linear models apply to mixed models as long as the predictor is a group level variable. Since in our case, the observations were grouped within-participants, numeracy was a group level variable, and so this method applies. The mediation coefficient was calculated by multiplying the numeracy $\Rightarrow$ conversions path by the numeracy $\Rightarrow$ Framing Effect path ($0.027 \times 0.07 = 0.0019$). Given the small sample size, we bootstrapped our estimate in order to simulate
variability in the coefficient and create an inferential statistic to test for significance. 10,000 resamples yielded 8 models that did not converge (likely because they did not include enough errors on the FOI), and the rest yielded a bias-corrected estimate equivalent to the original estimate and significantly greater than zero (p<.05 using normal and bias corrected tests). Therefore, we conclude that there is evidence of mediation, supporting Hypothesis 4.

*Between-Subject Framing*

To test between-subject framing effects, we used a crossed random effects model with judgments made only in the first frame encountered by each participant as the dependent variable, and frame, coded as a dummy variable, as a fixed predictor. The original study (Peters et al., 2006) used a larger sample (N = 100), but used 5 repeated measures that were highly correlated; all were judgments of student performance answering different percentages of exam questions correctly. Our study had a smaller sample (N=79), but employed more variated repeated measures (see design). Therefore, a direct power analysis could not be conducted, but we expected more power compared to the original study if the different scenario categories all behave similarly to the student scenario category.

We examined effects of numeracy, intelligence proxies, and foil of interest accuracy after the first frame, on the framing effect. We also examined the effects of numeracy and the intelligence proxies on answering the foil of interest correctly. We examined all 12 scenarios participants answered in the first frame they encountered for analyzing Hypothesis 1. However, Hypotheses 2 and 3 addressed the foil of interest, so
only those 4 scenarios that were followed by a foil of interest were used to address these hypotheses.

**The Framing Effect.** There was evidence for a framing effect such that responses in the positive frame (M=+1.11) were more positive than responses in the negative frame (M=-0.21), and this difference was significant (t = 6.72, t<.01). When examining the interaction between frame and numeracy, we found no effect. The less numerate made mean judgments of 3.59 and 4.56, and the more numerate 3.36 and 4.62 in the negative and positive frame respectively, with the interaction of numeracy and frame not approaching significance (t = -0.116, p = 0.91). The careful reader will notice that the direction of means is the opposite of the direction of the coefficient; this occurs because the coefficient reverses when controlling for intelligence proxies, but is essentially 0 in both cases, so this change does not warrant attention. This lack of support for Hypothesis 1 in a between-subject paradigm indicates a failure to replicate previous results (Peters et al., 2006). Unlike the original study, however, the present study examined several categories of scenarios: about students, snacks, milk and drugs. The original study only examined the frame by numeracy interaction for 5 student scenarios that were very similar to each other.

When we repeated the analysis separately for each category of scenario, the numeracy X frame interaction was significant only in the scenarios in the student category, and in fact the interactions for the other categories were in the opposite direction (not significantly), causing the non-significant trend in the interaction across scenarios. So it appears that out of the scenarios used in this study, the between-subject
framing effect is smaller for the more numerate only in the student scenario (these differences are summarized in Table 1). It is also possible that the results did not replicate because of the difference in methodology (i.e. memory trials, more repeated measures), however, this appears less likely since the Peters and colleagues’ results did replicate for the student scenario, and the order of scenarios was randomized across participants. Despite the significant two tailed t-test, this is a post-hoc explanation of the results, so we revisit this idea in a later study.

Another phenomenon perhaps worthy of note in this analysis is the effect of working memory on framing. It appears that working memory made the framing effect larger, and significantly so in two of the scenarios (see Table 1). It is not clear why this may have happened in this between-subject design. One possible explanation is that working memory allows people to maintain the exact wording of the scenario while making a judgment. In attribute framing, the exact wording is actually the source of the framing effect, and so those who can remember it might be expected to have larger effects. Fuzzy Trace Theory champions the idea that a gist of a stimulus is important, and that experts memorize the gist, which may be consistent with this idea.
<table>
<thead>
<tr>
<th></th>
<th>Drugs</th>
<th>Milk</th>
<th>Snacks</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>-0.61</td>
<td>-0.28</td>
<td>0.11</td>
<td>1.35</td>
</tr>
<tr>
<td>Numeracy</td>
<td>-0.92</td>
<td>-2.24*</td>
<td>-1.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Working Memory</td>
<td>-0.21</td>
<td>1.26</td>
<td>1.43</td>
<td>-0.15</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-0.04</td>
<td>-0.41</td>
<td>-1.38</td>
<td>-0.51</td>
</tr>
<tr>
<td>Frame X Num.</td>
<td>0.60</td>
<td>0.35</td>
<td>0.60</td>
<td>-1.96*</td>
</tr>
<tr>
<td>Frame X WM</td>
<td>2.18*</td>
<td>2.00*</td>
<td>1.40</td>
<td>0.91</td>
</tr>
<tr>
<td>Frame X Vocab</td>
<td>-0.73</td>
<td>-0.13</td>
<td>-0.72</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 1. Frame X Numeracy interaction performed separately for each item.

*Note.* These are results of multiple regressions with judgment on a scale from -3 to +3 as dependent variables and frame X individual difference measures interactions and main effects as independent variables. * = p<0.05.

**Foils Of Interest.** Hypothesis 2 predicted that incorrect responses on the foil of interest would be associated with smaller framing effects. To analyze this possibility, we examined a model with the judgment on a scale from -3 to +3 as the dependent variable and frame (-0.5 or +0.5), accuracy (1 or 0) on the foil of interest and their interaction as independent variables. The framing effect was non-significantly larger when participants answered the foil of interest correctly (responses to positive frame were 1.4 points higher than to the negative frame) than when they answered incorrectly (responses to positive...
frame were 0.72 higher than to negative frame), yielding a non-significant interaction of frame and correct vs. incorrect responses to the foil of interest (t=1.19, p>.2) and providing only weak directional evidence for Hypothesis 2. Unfortunately I could not examine this interaction for just the student scenarios because for those scenarios, none of the participants made errors on foils of interest in the positive frame of the student scenario (that is none of the 37 trials resulted in errors; see Table 2).

<table>
<thead>
<tr>
<th>Category</th>
<th>Negative Frame (n = 37)</th>
<th>Positive Frame (n=37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>92</td>
<td>95</td>
</tr>
<tr>
<td>Milk</td>
<td>84</td>
<td>92</td>
</tr>
<tr>
<td>Snack</td>
<td>89</td>
<td>95</td>
</tr>
<tr>
<td>Student</td>
<td>78</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2. Percent of participants answering to the foil of interest correctly (“did not appear”).

There was no support for Hypothesis 3 in the between-subject paradigm: numeracy did not predict responses to the foil of interest (z= 0.21 , p=0.83) in a crossed random effects logistic model. Again, this analysis could not be repeated for only the student scenarios because the foil of interest was always answered correctly.

The between-subject analyses provided only limited support for Hypotheses 1-3, not reaching conventional levels of significance, so we did not test for the mediation hypothesis. This lack of an effect could be due to a small sample or to the difference in
mechanisms (e.g., memory effects, effects of thoughtfulness) between within- and between-subject framing effects discussed in the introduction. The effects reported are when controlling for intelligence proxies.

Discussion

The present experiment provides evidence that the framing effect is attenuated when the alternative frame is brought to mind with a conversion. It also suggests that these conversions are in part responsible for the smaller framing effects that the more numerate display. Finally, it appears that this mechanism may only be present when the framing manipulation is done within-subjects. The discussion below explores the within-subject results. The comparable between-subject effects are addressed in Study 2 and later discussions.

Adding to previous research, this study provides support for the existence of a mechanism used by the more numerate. It also provides a possible way to aid the less numerate. Future research may show that advising the less numerate to consider alternative formulations of a scenario may help them make conversions that the more numerate make autonomously.

The study also adds to the attribute framing literature. It provides evidence that the same individual difference measure – numeracy, and the same thought process – a numeric conversion, can help ameliorate attribute framing effects across a range of dissimilar problems, implying that the same mechanism is at work across the various instances of attribute framing. However, it also suggests that different processes may be operating in within- and between-subject framing effects. That is, the two manipulations
cannot be used interchangeably as different instances of the same effect. It appears that in the within-subject effect, participants are more likely to compare the frame to the previously encountered frame.

Follow up research should disentangle the direction of the causal relationship between numeric conversions and the conversions of the framed information. The numeric conversions might be a product of comparing a scenario in the second part of the experiment to its original form (i.e., this mathematical operation is necessary for my consistent judgment, and since I am highly numerate I can do it). Alternatively comparing a scenario to its original form might be the product of the conversions (i.e., I do a mathematical operation just because I am highly numerate and as a result I get something that is useful for consistency in my judgment). Cokely and Kelley (2009) would appear to argue for the latter explanation, but at this point no experimental evidence exists.

The study also leaves for future investigation the question of what exactly it is about numerate people that elicits these conversions. While one might assume the “effective ingredient” of numeracy is numeric skill, given the simplicity of these conversions, it is likely that even the least numerate among college students have what it takes to make the conversions. Attention to numbers or propensity to think about them may be correlated to numeric skill and may instead be the factor that drives the relationship between numeracy and these conversions. In addition, since the conversions do not fully mediate the relation between numeracy and better decision making with regard to framed information, other mechanisms may be in play, or the rest of the effect
of numeracy on the framing effect could be accounted for using a cleaner measure of conversions than errors on the foil of interest.

We can speculate about some of these mechanisms that differentiate the more numerate from the less numerate and explain varying decision making outcomes. For example, it has been shown that numerate individuals are able to discriminate more precisely between numbers than less numerate individuals (e.g., Peters et al., 2009). The objective stimulus “85%” may have a subjective representation that can be thought of as a distribution with mean at 85 and variance that depends on a number of factors including one’s numeracy. If an individual has an extremely wide distribution, telling him “John got 85% questions correct” may be the same as telling him “John got some questions correct,” since he is unable to discriminate between numbers well. Alternatively, telling him “John got 15% questions incorrect” translates to “John got some questions incorrect.” When comparing “John got some questions correct” and “John got some questions incorrect,” the former option is clearly superior because it excludes the possibility of “John got no items correct,” whereas the latter excludes the possibility of “John got all items correct,” but the term some is so vague that neither provides any information beyond this. Therefore, the numerate may also make more balanced judgments due to more precise numeric representations.

Another possibility is that the difference is attentional. The numerate may pay attention to numeric stimuli rather than verbal ones when looking for information in order to make a decision. Effectively, this mechanism results in a situation similar to the previous one: the more numerate concentrate on exactly what percentage of questions
John answered correctly, while the less numerate concentrate on the fact that John answered some questions incorrectly.

Study 2

The between-subject results from Study 1 were not very encouraging. Though we did find a framing effect, we failed to replicate the numeracy by framing effect interaction, except in the same scenario category (about students) that Peters and colleagues (2006) used. However, this by-scenario analysis between-subjects was not planned and the study may have been underpowered for such an analysis.

In Study 2, we collected a large sample in order to be able to examine the framing effect across levels of numeracy for different framing scenarios. We hoped that the differences found in Study 1 were spurious (since they resulted from a post hoc analysis) and that the framing effect by numeracy interaction was robust to variation in framing scenario and to controls for intelligence. In this study, we collected a more direct measure of fluid intelligence (as opposed to working memory, vocabulary or test scores), a modified version of Raven’s Matrices. In addition, we wanted to check whether the original results are robust to percentages that were round (e.g., 80%) and precise (e.g., 79%). We expected that the hypothesized conversions will be more likely if the percentages are easier to convert because they are round and therefore the framing effect would be smaller.
Hypothesis 5:

The framing effect will be smaller when a percentage is easy to convert (e.g. 70% to 30%) than when a percentage is difficult to convert (e.g. 68% to 32%).

We also hoped to rectify one concern with recognition errors. They may result simply from associations between complementary percentages that the more numerate have, not from online conversions. That is, the numerate may have made more errors because they simply associate 75% with 25%, not because they made a conversion specific to the framing problem. The fact that the errors mediated the relation between numeracy and the framing effect, makes this explanation less likely, but mediation models are far from causal, so it is preferable to find experimental evidence. In addition, the very act of collecting recognition data may have changed the way participants processed subsequent framing scenarios. In order to rule out this possibility and test our within-subject findings in a between-subject design, we assessed recall errors. If the more numerate recall 25% instead of 75%, it cannot be because 25% was activated due to its association with 75% (since 75% would be activated more strongly, and therefore would be the first to be recalled). A surprise recall task at the end of the experiment was employed. Therefore Study 2 was intended to address the hypotheses of Study 1 using a new measure of conversions and a larger sample size.

A useful byproduct of this experiment is that by varying the percentage and frame in a crossed fashion (i.e., manipulating the size of the percentage in each frame for the same scenario between-subjects), we can actually estimate the magnitude of the framing effect in objective percentage points rather than simply in subjective evaluations. From a
large number of experiments on attribute framing and from Study 1, we know changing a frame from a negative to a positive frame would increase the evaluation of a student, drug, snack or milk product by some amount on an objective scale (such as the scale from -3 to +3 used in Study 1). However, the scale is arbitrary and its representation is probably not linear. In this experiment, we can estimate the size of the framing effect in terms of percentages of the attribute. For example by using the coefficient of percent size and of frame, we may conclude that 90% of the questions correct is subjectively equivalent to 20% of the questions incorrect (despite the fact that it is normatively equivalent to 10% of the questions incorrect).

**Method**

**Participants**

Data from 359 participants (Mean age = 19, 55% female) were collected in exchange for course credit. Some participants’ data (n=17) were lost due to software problems or experimenter error. Another 5 participants’ data were deleted because they spoke English for less than 4 years, yielding a total of 337 participants whose data were analyzed, over 3 times larger than the sample of Peters et al. (2006). To the best of our knowledge, these data can be considered missing at random, since none of the sources of missingness were correlated to experimental variables, except perhaps the lack of knowledge of English. Elimination of the participants who did not speak English well could be justified because we are interested in an English speaking population.
Design

Each participant saw 4 framing scenarios, from Study 1, except they included less irrelevant detail (irrelevant detail was included in Study 1 to make the recognition task reasonably difficult). After several unrelated tasks, the participants were asked to recall a part of the framing scenario that included the percentage. For example, if the original stimulus was “Emily took a psychology test. She answered 87 % of the questions correctly. Please rate this student's performance compared to other students,” the recall question would ask participants to fill in the missing part in "Emily took a psychology test. She answered _________________. Please rate this student's performance compared to other students.”

The percentages that were embedded in the scenarios were either easy or difficult to convert. The round (and therefore hypothetically easy to convert) percentages were 90-10 and 60-40. In a pilot study, trials in which participants were required to subtract each of these numbers from 100 were among those solved most quickly and therefore, presumably, most easily. The precise (and hypothetically difficult to convert) percentages were 87-13 and 69-31; these were reasonably close to the round percentages and were some of the most difficult to subtract from 100 in the same pilot study. In the design, we crossed the percentages with the framing scenario and with frame, for a 4 (framing scenarios, within-subject) x 4 (percentage pairs, between-subject) x2 (frame, between-subject) design. The frame condition was chosen for each participant in the beginning of the study, so no participant saw the positive frame of one stimulus and the negative frame of another. The percentage within the pair was decided by the frame, so no extra factor
was needed for that (e.g., they were presented student scores that were 90, 60 87 or 69% correct, or 10, 40 13 or 31% incorrect).

In order to get a good measure of the framing effect, and to make sure that our findings are robust to manipulations exogenous to our theory, it is necessary to use multiple types of scenarios, as had been done in Study 1. Since we were manipulating the difficulty of the conversion through the percentage in the problem, we made sure that the association between percentage and the problem was random, unlike in Study 1, where the percentage was not directly addressed and all percentages were precise. If this precaution had not been taken, the content of the problem would be confounded with Hypothesis 5. That is, we randomized the 4 percentages used with the 4 categories of scenarios (student, milk, snack and drug) that were used and the 2 frames, for a total of 32 scenarios. Each participant viewed 4 of these 32 resulting scenarios, such that they were all of different categories, and all 4 were of the same frame, but they were of different percentages.

The short version of the numeracy scale was used in this experiment (Weller et al., 2012). Participants also completed a modified Raven’s Matrix fluid intelligence measure to control for intelligence. Standardized test scores were collected as well. Only a small portion of the sample took the SAT and the ACT did not differentiate between math and English and science scores, so these were not used in the analysis. Pairwise correlation matrices and scatterplots as well as nonlinear fits of all individual difference measures for studies 1-3 are available in Appendix E.
Results

Framing Effect. Model 0 was fit in order to check if we replicated the attribute framing effect and how large this framing effect was. Model 0 was a mixed model with judgments (i.e. responses to the attribute framing scenarios on scales from -3 to +3) as a dependent variable, a random intercept and slope for the different scenarios and a random intercept and slope for the participants\(^2\), and fixed effects for frame and percentage size. There were 337 (subjects) X 4 (scenarios per subject) = 1348 observations. The effect of frame was 1.0 point on the scale from -3 to +3 (t=3.86, p<.01). The effect of percentage size was 0.05 (t = 5.52, p < 0.01; see Table 3). Comparing these effects\(^3\) showed that the framing effect was on average equivalent to increasing the percentage by 20 points (though this number differs across scenarios). That is according to our model, “Emily answered 70% correctly” would be judged approximately equally to “Emily answered 10% incorrectly,” a substantial effect.

|                | Estimate | S.E. | T ; p(|t|) |
|----------------|----------|------|-----------|
| Intercept      | -0.53    | 0.87 | -0.61 (p=0.54) |
| Frame          | 0.05     | 0.01 | 5.52 (p<0.01) |
| Percentage Size| 1.00     | 0.26 | 3.86 (p<0.01) |

Table 3. Unstandardized regression coefficients of Model 0.
We fit Models 1 through 4 (see Table 4) to examine Hypothesis 1 (that the framing effect is smaller for the more vs. less numerate). Model 1a was an expanded version of model 0 that also included frame (coded as -0.5, +0.5), and numeracy, and their interaction. We also fit Model 1b in order to examine the possibility that the results of numeracy in Model 1a were due to correlation with intelligence. This was an expanded version of Model 1a, that included all of Model 1a’s terms but also included a Raven’s Matrices score and an interaction of Raven’s Matrices with frame. There is a nonsignificant relationship in the expected direction when examining the interaction of numeracy with frame in Models 1a and 1b (t=-1.08, p = 0.28, and t=-0.92, p = 0.36, respectively).

Since the effects were not significant, and as planned from the findings in Study 1, we examined the scenario about student performance separately in Models 2a and 2b. Models 2a and 2b were like Models 1a and 1b, except that they were fit to a subset of the data, that is, the student scenario. Since these models had no repeated measures, we used a robust linear model (iteratively re-weighted least squares; Huber, 1981) rather than the hierarchical models used in Models 1a and b. Robust models are just like multiple regression, except they underweight overly influential observations. Models 2a and b examined 337 (subjects) X 1 (scenario per subject) = 337 observations. The expected interaction between frame and numeracy was absent (t=-1.03, p = 0.30; see Table 4 for full results). Altogether, Study 2 did not appear to support Hypothesis 1. We did replicate Peters and colleagues’ (2006) findings and our own findings from Study 1 conceptually, but only when rerunning model 2a with a combination of the numeracy scale and the
math portion of SAT instead of just numeracy to predict judgments for the student scenario \((t = 2.0, p = 0.05)\). It suggests that the effect of mathematical skill on framing for the student scenario may be due to a broader array of numeric skills than that measured by the 8 item Weller et al. (2012) numeracy scale or simply that the scale is too noisy.

It was surprising that we did not significantly replicate our own results on the student question from Study 1 or Peters and colleagues’ (2006) results with the numeracy scale despite the large sample size, though we did only have one scenario about students and Peters and colleagues and Study 1 used repeated measures. We looked at the differences between stimuli in the present study and the originals. We noticed that their scenarios and our scenarios from Study 1 included only precise percentages, whereas our scenarios in Study 2 included both precise and round percentages (though the effect presumably should not depend on whether the percentage is round).

To have an empirical basis to analyze the effects of numeracy on framing for round and precise percentages separately, we needed to find a third order interaction between numeracy, frame and whether the percentage was round or precise. Therefore we fit Model 3 (the results of which are reported in Table 5), another robust linear model, which used the main effects and all possible interactions of frame (coded -0.5, +0.5), numeracy and roundness of the percentage (coded 0 if the percentage was 87, 69, 13 or 31 and coded 1 if the percentage was 90, 60 10 or 40) to predict judgments. We did find the third-order interaction when using a one tailed test, which was appropriate since we were checking for a direct replication of a previous experiment \((t = 1.61, p = 0.05, \text{ one tailed})\). This interaction was in the expected direction, such that the percentages that were
precise (like in Peters and colleagues’ (2006) original experiment) exhibited the frame by numeracy interaction, while the round ones did not (see Figure 3 for a bar graph of the means).

Since the interaction was significant, we repeated Models 2a and 2b separately for the scenarios with round percentages and precise percentages. Models 4a and 4b are the analogues of Models 2a and 2b for only precise percentages, and Models 5a and 5b are the analogues for round percentages. It is important to notice that the numeracy by frame interaction is significant with precise numbers in Models 4a and 4b, when one tailed tests are used (which are again appropriate because this is a direct replication), but non-significantly in the opposite direction with round numbers in Models 5a and 5b (see Table 4, and see Study 3 where this effect is stronger). This appears to be strong evidence that whether a percentage is round or precise has a strong impact on the interaction of the framing effect and numeracy. Again, this interaction is pictured in Figure 3.

These findings are inconsistent, however, with the explanation that conversions are the only factor that explains the numeracy X frame interaction in the student scenario. If it were only the conversion, we would expect one of three possibilities: (1) Everybody makes conversions in the round condition, but nobody makes conversions in the precise condition, in which case there should be no numeracy X frame X round interaction; (2) Everybody makes conversions in the round condition, but only the numerate make conversions in the precise condition, in which case everybody should have smaller framing effects in the round condition, but only the numerate should have smaller effects in the precise condition; or (3) Only the numerate make conversions in the round
condition, and nobody makes conversions in the precise condition, in which case
everybody should have large framing effects in the precise condition, and only the
numerate should have small framing effects in the round condition. The results support
none of these possibilities. In fact, what appears to be the case in Figure 3, is that the
more numerate have smaller framing effects only in the precise condition, and in the
round condition both groups have the same, relatively large, framing effects (with the
more numerate trending to have larger effects).

It therefore appears that roundness of the percentage has some non-intuitive effect
on the numeracy X frame interaction in the student scenario. One possible explanation is
that the conversions have an effect on framing only if they are deliberated on sufficiently.
In the round condition, everybody makes the conversions, but they are very easy for the
more numerate, so the more numerate do not deliberate about them sufficiently. On the
other hand in the precise condition, only the more numerate make the conversions, and
since these are relatively difficult, the more numerate have to deliberate.
<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>Intercept</th>
<th>Percent Size</th>
<th>Frame Numeracy</th>
<th>Raven X Num.</th>
<th>Frame X Raven</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Entire Sample</td>
<td>-0.55</td>
<td>5.52</td>
<td>3.67</td>
<td>-0.27</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=0.58</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>p=0.79</td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>Entire Sample</td>
<td>-0.60</td>
<td>5.54</td>
<td>3.41</td>
<td>-0.18</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=0.55</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>p&lt;0.86</td>
<td>p&lt;0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.92</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Table 4. Crossed random effects models describing effects of numeracy on framing.
<table>
<thead>
<tr>
<th>Model</th>
<th>Sample</th>
<th>Intercept</th>
<th>Percent Size</th>
<th>Frame</th>
<th>Numeracy</th>
<th>Raven</th>
<th>Frame X Num.</th>
<th>Frame X Raven</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a</td>
<td>Student Scenario</td>
<td>2.39</td>
<td>6.49</td>
<td>2.57</td>
<td>0.76</td>
<td>---</td>
<td>-1.03</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=0.02</td>
<td>p&lt;0.01</td>
<td>p=0.01</td>
<td>p=0.45</td>
<td></td>
<td>p=0.30</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>Student Scenario</td>
<td>-2.53</td>
<td>6.53</td>
<td>2.73</td>
<td>0.39</td>
<td>0.88</td>
<td>-0.63</td>
<td>-0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
<td>p=0.01</td>
<td>p=0.70</td>
<td></td>
<td>p=0.53</td>
<td>p=0.35</td>
</tr>
<tr>
<td>4a</td>
<td>Student Scenario</td>
<td>-1.17</td>
<td>2.41</td>
<td>2.70</td>
<td>2.66</td>
<td>---</td>
<td>-1.97</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Precise %</td>
<td>p=0.24</td>
<td>p&lt;0.02</td>
<td>p=0.01</td>
<td>p=0.01</td>
<td></td>
<td>p=0.03, one tail</td>
<td></td>
</tr>
<tr>
<td>4b</td>
<td>Student Scenario</td>
<td>-1.43</td>
<td>2.58</td>
<td>2.93</td>
<td>2.45</td>
<td>0.99</td>
<td>-1.69</td>
<td>-1.02</td>
</tr>
<tr>
<td></td>
<td>Precise %</td>
<td>p=0.15</td>
<td>p=0.01</td>
<td>p&lt;0.01</td>
<td>p&lt;0.02</td>
<td></td>
<td>p=0.05, one tail</td>
<td>p=0.31</td>
</tr>
<tr>
<td>5a</td>
<td>Student Scenario</td>
<td>-1.89</td>
<td>6.35</td>
<td>1.31</td>
<td>-1.17</td>
<td>---</td>
<td>0.23</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Round %</td>
<td>p=0.06</td>
<td>p&lt;0.01</td>
<td>p=0.19</td>
<td>p=0.24</td>
<td></td>
<td>p=0.82</td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>Student Scenario</td>
<td>-2.00</td>
<td>6.34</td>
<td>1.40</td>
<td>-1.52</td>
<td>0.91</td>
<td>0.43</td>
<td>-0.52</td>
</tr>
<tr>
<td></td>
<td>Round %</td>
<td>p=0.05</td>
<td>p&lt;0.01</td>
<td>p=0.16</td>
<td>p=0.13</td>
<td></td>
<td>p=0.67</td>
<td>p=0.60</td>
</tr>
</tbody>
</table>

Table 5. Results of Models 2-5

*Note.* This table displays t-values of all the effects, with corresponding p-values. Each model has two lines (a and b), one examining the interaction of frame and numeracy, and
the other examining the same relationship while controlling for the effects of intelligence (measured using Raven’s Matrices). Judgment is the dependent variable in all models. Model 3, which was run on precise percentages, gets significant interaction between numeracy and frame in the expected direction, while the trend in Model 4, which is run on round percentages is in the opposite direction. Also note that the percentage size was included as a control.

| Coefficient               | T   | P(>|t|) | One Tailed? |
|---------------------------|-----|--------|-------------|
| Percentage Size           | -3.55 | <0.01  | No          |
| Numeracy                  | 2.78 | 0.05   | No          |
| Frame                     | 2.83 | <0.01  | No          |
| Percentage Round          | 2.80 | 0.01   | No          |
| Numeracy X Frame          | -2.06 | 0.02   | Yes         |
| Numeracy X Round          | -2.83 | 0.02   | No          |
| Frame X Round             | -1.12 | 0.27   | No          |
| Numeracy X Frame X Round  | 1.61 | 0.05   | Yes         |

Table 6. Results of Model 3.

Note. Tests on numeracy were conducted as one tailed tests because they were direct replications of previous experiments.
Figure 3. Means for data corresponding to models Model 3, 4 and 5 (all for the student question).

*Note.* The panel on the right corresponds to scenarios which included percentages that were round: 90, 60, 10 and 40. The panel on the left corresponds to scenarios in which the percentages were precise: 87, 69, 13 or 31. Median split of 5 was used to generate the more and less numerate groups. The interaction in the left panel, but not the right panel is significant, indicating that when percentages are precise, greater numeracy is associated with a smaller framing effect. However, when percentages are round, there is a trend in the opposite direction.

**Conversions.** In this experiment, we coded a conversion whenever a participant recalled the percentage that corresponded to the frame s/he was NOT presented. For example, if presented “Emily answered 90% correctly” and when asked to fill in the blank in “Emily answered ________” and the participant recalled anything that involved 10%, the response was counted as a conversion.
Next, we examined the independent effects of whether there was evidence for a conversion, and the percentage was easy or hard to convert (i.e., round or precise respectively) on the framing effect. We expected framing effects to be smaller when conversions were made above and beyond the effects of roundness. Almost every recalled converted percentage in this study was round (that is, all but one of the memory errors consistent with conversions occurred when the percentage was precise; see Table 7). To determine whether this was due to chance, we fit Model 6, a logistic regression that uses recall in a manner consistent with conversions as the dependent variable (coded 1 if and only if the participant recalled the converted percentage, and coded 0 if the participant recalled correctly or made some other recall error; with 337 subjects X 4 memory trials per subject = 1348 observations), and used roundness (coded 1 if the percentage in the stimulus to be recalled was round and 0 if it was precise) as the independent variable. There was an effect such that round percentages resulted in recall errors consistent with conversions more often than did precise percentages (z=3.3, p<.01). This may be a function of the measurement of conversions (both round and precise numbers may be converted, but only round conversions may be recalled), or a genuine effect (only round numbers are converted and therefore only round conversions are recalled), and so it is unclear what theoretical implications this finding has. However, it made it necessary to control for the ease of conversion when measuring the unique effect of conversions on other variables and vice versa in the following analyses.
Hypothesis 2. (When participants made conversions, their between-subjects framing effects would be smaller than when they did not). To test Hypothesis 2, we fit Model 7, a crossed random effects model with random intercepts and slopes for participants and scenarios and with judgment as the dependent variable (with all 1348 observations) and frame, recall error (coded as in Model 6) and their interaction, as independent variables. Roundness (coded as in Model 3) and its interaction with frame were included as covariates for reasons outlined in the previous paragraph. The size of the percentage was also included as a covariate. The effect of the interaction between frame and conversion was significant such that the framing effect was smaller when conversions were made ($t=-2.13$, $p = 0.03$; see Table 7 for complete results). Since the interaction coefficient is larger than the effect of frame, we see that those people who recalled the alternative frame had no framing effect (in fact it was the opposite direction).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Number of conversions when % is Precise</th>
<th>Number of conversions when % is Round</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snack</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Drug</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Milk</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Student</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 7. Number of memory errors consistent with conversions
Hypothesis 2 could not be tested for the student scenario separately, because no errors were made when the information was presented in the positive frame.

|                                | Estimate | T  | p(>|t|) |
|--------------------------------|----------|----|--------|
| Percentage Size                | 0.05     | 5.69 | <0.01  |
| Frame                          | 0.88     | 7.96 | <0.01  |
| Percentage Round               | 0.05     | 2.86 | <0.01  |
| Recall Error Consistent with Conversion | 0.67     | 0.44 | 0.67   |
| Frame X Round                  | 0.30     | 2.10 | 0.04   |
| Frame X Conversion             | -1.11    | -2.13| 0.03   |

Table 8. Results of Model 7.

*Note.* Judgment was the dependent variable. Percentage Round is a dummy variable = 1 when the percentage in the scenario judged was round and = 0 when it was precise. Recall Error Consistent with Conversion is a dummy variable = 1 when the participant made a recall error consistent with a conversion in the recall trial that corresponded the scenario judged, 0 otherwise.

**Hypothesis 3** (that the more numerate tended to make more errors consistent with conversions) was tested using a mixed effects logistic model with recall errors as the dependent variable (again, coded as in Model 6). Model 8 had random intercepts for subjects and scenarios, and numeracy and Raven score as predictors. There was no significant effect of numeracy (z=0.76, p = 0.45) or Raven’s Matrices (z = -0.08, p =
In Model 9, we went on to examined the same dependent and independent variables as in Model 8, but we used data only from the scenario about student performance (337 (subjects) X 1 (scenarios per subject) = 337 observations). Since there were no repeated measures here, Model 8 was a logistic model. A unique effect of numeracy was present, such that the more numerate made more errors consistent with Hypothesis 3 (z=2.2, p = 0.03), while those who scored higher on Raven’s Matrices tended to make fewer errors (z = -1.5, p = 0.14). These results indicated that the numerate (not the intelligent) were more likely to make conversions, but only in the student scenario.

**Hypothesis 5.** The idea that placing easy-to-convert percentages in framing scenarios would result in smaller framing effects than difficult-to-convert percentages, received no support. To test this, we examined the frame by roundness interaction in Model 7 (see Table 7). Contrary to what we originally expected, scenarios that had percentages that were easy to convert had larger framing effects regardless of whether conversions were controlled for or not (t=2.1, p = 0.03 in Model 7; t=1.9, p=.05 in another model that was equivalent to Model 7, except it did not control for conversions).

**Discussion**

The results of this experiment revealed some unexpected trends. First, the framing effect X numeracy interaction between-subjects is not the same across categories of scenarios, round vs precise percentages in those scenarios, and also possibly across different measures of numerical skill. To resolve these sorts of issues, we would need a
large pool of framing items and a large pool of mathematical competence and numeracy items.

Across scenarios, it appears that making percentages round (e.g., 90% vs 89%) had the intended effect of making conversions more likely, or at least of making it more likely to observe error responses resembling conversion. However, above and beyond causing conversions, round numbers had the unexpected effect across scenarios of being associated with larger framing effect (though there was a trend for the more numerate to show the expected effect). A post-hoc explanation is that precise numbers had larger framing effects because precise percentages are less fluent and therefore cause more elaborative processing (Alter, Oppenheimer, Epley & Eyre, 2007). Investigating whether this is the case is not the primary purpose of these studies, but might make for interesting further work.

Some hypotheses were supported. When participants made recall errors consistent with conversions, the framing effect was smaller. Perhaps the most interesting conclusion from Study 2 is that people do, in fact, independently and without instruction, convert framed information, even without having encountered the alternative frame in the experiment. Such conversions appear to nullify the framing effect. It was not clear that this was the case in Study 1.

Interestingly, recollections of the alternative frame were about the same across levels of numeracy when analyzing the entire dataset (Model 7). However, for the student question, the more numerate did make more responses consistent with conversions (Model 8).
It is interesting and inconsistent with the original hypotheses that the different framing scenarios appear to behave differently with respect to numeracy. For example numeracy predicts the framing effect for the student scenario, but not across all 4. Numeracy also makes conversions more likely in the student scenario, but not across all 4. It is not immediately clear what features of these scenarios moderate what mechanisms. We attempted to examine some differences in unreported analyses with little success.

Perhaps a more fruitful approach would be to attempt to gather information about the most commonly used framing scenario, to inform previous and future research on attribute framing. It is likely that the framing scenario most often used is the original Levin & Gaeth (1988) scenario about a beef product, in which participants are told they are to cook lasagna for a relative with beef that is 25% fat or 75% lean. The experiment was originally run with real cooked beef that was tasted by participants, but the results have since been replicated a number of times with a number of variations, and usually with a hypothetical judgment (e.g., Donovan & Jalleh, 1999; Johnson, 1987; Keren, 2007; Janiszewski, Silk & Cooke, 2003). Participants are usually asked to appraise the fat, greasiness and overall quality of beef. Since this scenario is so often used, we wanted to test whether Hypotheses 1 through 4 would receive support in this framing scenario. This design would also get rid of any within-subject repetition or learning effects associated with seeing several attribute framing scenarios.
Study 3

This study aimed to test Hypotheses 1-4 in a between-subject design using the same methodology as Study 1 (that is, conversions were again operationalized as recognition errors). It was designed and run simultaneously with Study 2, and therefore we could not use what we learned about round and precise percentages, or expand further on those conclusions. Since Study 1 appeared underpowered to make conclusions about the between-subject framing effect, Study 3 was conducted with a large sample size. In other words, this study was intended to replicate the conclusions of Study 1 in a between-subject design and on a scenario commonly used in the literature.

Therefore, this study used a different attribute framing scenario than any of those used in either Study 1 or Study 2. Study 3 did not use repeated measures, and so was free of any order or repetition effects that may have confounded the previous two studies. This study used the same participants as Study 2, but was run first. Since Study 2 employed repeated measures within its own design, we do not expect that preceding it with an extra scenario caused any problems.

Finally, each participant only saw one percentage foil. That is, following the framing scenario, all participants went on to view a number of recognition stimuli as in Study 1. However, some participants eventually saw the foil of interest (i.e., 20%), and others eventually saw another percentage foil that was similar to the foil of interest in quantity and in roundness, but did not correspond to the converted percentage in the original scenario (i.e., 10%). The study was designed to allow us to use an interaction
between the type of foil and whether it was answered correctly to predict the framing effect.

Method

This study was run at the same time as Study 2 on the same 359 participants. The reader is therefore directed to examine Study to learn how the participants were recruited and why the data from some were lost or deleted.

The method of this study was almost exactly the same as that of Study 1, except participants viewed only one framing scenario after viewing the instructions. This scenario was as follows (bold added for emphasis):

You are inviting a special friend to dinner, and you have decided to make their favorite lasagna dish with ground beef. The beef you are using is **80% lean (or 20% fat)**.

They were then asked to make a judgment about how fat or lean the meat is, how greasy or greaseless the meat is, and finally a judgment of the overall quality of the meat. On a scale from -3 to +3 that was labeled appropriately for the particular judgment. After these 3 evaluations they proceeded to the recognition phase.

Unlike in Study 1, the recognition trials were not randomized. Randomization adds noise, but was necessary in Study 1 so that participants did not learn when to expect the percentage targets and foils. Instead, the participants viewed 8 distractor foils, after which some participants viewed the converted percentage, that is the foil of interest (20% in the positive frame or 80% in the negative frame, the foil of interest), and others viewed another percentage that neither appeared in the scenario nor was the complement of the percentage that appeared in the scenario, but superficially matched the foil of interest
(10% in the positive frame or 90% in the negative frame). Finally on the last recognition trial, they viewed the correct percentage (80% in the positive frame or 20% in the negative frame). We specifically chose the foil percentages to appear before the correct percentages in order to make errors more likely (they were somewhat rare in Study 1, and rare errors are statistically harder to model than common errors).

Since there were no repeated measures, there was some worry that extreme (influential) observations for some participants may dramatically affect the results. The data were therefore analyzed with a robust linear model which weights influential observations less heavily. However, the results were similar to regular multiple regression.

**Results**

**Variables of Interest.** The three measures of fat, greasiness and overall quality of beef were not as highly correlated as expected, but correlated acceptably (fat and greasiness, 0.54; fat and quality, 0.58; grease and quality, 0.44; alpha = 0.77). We therefore calculated factor scores using Thurstone’s regression method and refer to these values as “factor values” in the rest of the analysis. Factor values can be thought of as a z-distributed representation of the factor corresponding to how good a participant thought the beef was, compared to other participants.

Another variable that was important to the following analysis was whether or not each the foil that was presented was answered incorrectly. We expected those participants who saw a foil of interest (20% in the positive frame or 80% in the negative frame, in this study) to err significantly more than those who saw another foil (10% in the positive
frame, or 90% in the negative frame). This effect was not observed. In our sample, 75% of participants answered the foil of interest correctly, while 70% of participants answered the other foil correctly (z=1.0, p =0.3). For comparison, 95% of participants answered the target percentage correctly.

**The Framing Effect.** We fit a robust linear regression with factor values as the dependent variable and frame (coded -0.5, +0.5) as the independent variable and observed a between-subject framing effect of 0.97 on the z-distributed factor values (t=17.2, p<0.05), indicating that factor values in the positive frame were almost a full standard deviation higher than factor values in the negative frame. This result replicated the large number of papers written on the attribute framing effect for this scenario.

**Hypothesis 1** (that the framing effect is smaller for the more numerate). We then fit Models 9a and 9b, with factor values as the dependent variable, and frame, numeracy and their interaction as independent variables in Model 9a. Model 9b was the same as Model 9a, but also included Raven’s Matrix scores and their interaction with frame. Surprisingly, the framing effect tended to be larger for the more numerate (t=1.84, p = 0.07, in Model 9a), and this relation was reduced when controlling for general intelligence (t=1.43, p = 0.15, in Model 9b; see Table 8 for full results of Model 9b). Therefore Study 3 provided some evidence against Hypothesis 1.
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th></th>
<th>T-Value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>0.57</td>
<td>0.26</td>
<td>2.20</td>
<td>0.03</td>
</tr>
<tr>
<td>Numeracy</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Raven</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.78</td>
<td>0.44</td>
</tr>
<tr>
<td>Frame X Numeracy</td>
<td>0.07</td>
<td>0.05</td>
<td>1.43</td>
<td>0.15</td>
</tr>
<tr>
<td>Frame X Raven</td>
<td>0.03</td>
<td>0.04</td>
<td>0.66</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 9. Results of Model 9b.

*Note.* Numeracy appears to have the opposite of the expected effect on framing given results of Peters et al (2006). However this is not unexpected given that in Study 2 we found the same reversal when percentages were round.

**Hypothesis 2** (that conversions predict the framing effect). We expected a third order interaction between frame, whether the recognition foil was answered correctly, and whether the recognition foil was a foil of interest, or just a percentage foil in predicting judgments. Specifically, we expected the framing effect to be smaller when the foil of interest was answered incorrectly (indicating a conversion) but not when another foil was answered incorrectly (giving no information about a conversion). To test this interaction we fit Model 10, another robust linear model, with the factor value as the dependent variable, and frame, type of foil (dummy variable that was 1 when the foil was a foil of interest, and 0 when it was another foil) and correct rejection (dummy variable
that was 1 when the foil was answered correctly and 0 when answered incorrectly) and all
the possible interactions of these three variables. No third order interaction was found (t =
0.45, p = 0.65; see Table 9 for full results), and in fact only the simple effect of frame
emerged as significant.

|                  | Estimate | S.E. | T-Value | P(>|t| ) |
|------------------|----------|------|---------|----------|
| Frame            | 1.03     | 0.22 | 4.69    | <0.01    |
| Correct ( = 1 when foil is rejected) | 0.09 | 0.13 | 0.71    | 0.48     |
| Type of foil ( = 1 when FOI)         | -0.04    | 0.16 | -0.23   | 0.82     |
| Correct X Type of foil       | 0.00     | 0.19 | 0.00    | >0.99    |
| Frame X Correct          | 0.00     | 0.26 | 0.00    | >0.99    |
| Frame X Type of foil      | -0.23    | 0.32 | -0.70   | 0.48     |
| Frame X Correct X Type of foil | 0.17 | 0.38 | 0.45    | 0.65     |

Table 10. Results of Model 10.

Note. Correct is a dummy variable = 1 when the foil is rejected (the participant correctly
indicates it was not in the scenario) and = 0 when the foil is accepted (the participant
erroneously indicates it was in the scenario). No significant results to report.

**Hypothesis 3** (that the more numerate are more likely to make conversions). We
expected that the more numerate would be less accurate on foils of interest, but equally or
more accurate on other foils. Therefore, we fit Model 11, a logistic model with
recognition on the foil as the dependent variable, and the type of foil, numeracy and
Raven score, as well as interactions of numeracy and Raven score with the type of foil as
independent variables. We found that the more numerate do err more on foils of interest,
compared to other foils (t=-1.6, p=0.11), but this relation is partially accounted for by
general intelligence, which turned out to be a stronger predictor (t=-2.24, p=0.02; results
are summarized in Table 10 and means are in Table 11). Nonetheless, these results
suggested that people with higher ability, and perhaps the more numerate are more likely
to answer foils of interest incorrectly, while they are more likely to answer otherwise
identical percentage foils correctly, suggesting that their errors on foils of interest
indicate conversions.

|                           | Estimate | S.E. | z value | Pr(>|z|) |
|---------------------------|----------|------|---------|----------|
| Intercept                 | -0.70    | 0.58 | -1.20   | 0.23     |
| Foil of Interest (FOI)    | 2.91     | 0.83 | 3.50    | 0.00     |
| Numeracy                  | 0.09     | 0.10 | 0.86    | 0.39     |
| Raven Matrix Score        | 0.22     | 0.09 | 2.37    | 0.02     |
| FOI * Numeracy            | -0.24    | 0.15 | -1.60   | 0.11     |
| FOI * Raven               | -0.29    | 0.13 | -2.24   | 0.02     |

Table 11. Results of Model 11.

*Note.* Foil of Interest is a dummy variable = 1 when the foil is a foil of interest, and = 0
when it is another foil. The FOI*Numeracy interaction is significant when not accounting
for intelligence.
Table 12. Means of key variables in Model 11.

<table>
<thead>
<tr>
<th></th>
<th>Less Numerate</th>
<th>More Numerate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foils of Interest</td>
<td>0.81</td>
<td>0.65</td>
</tr>
<tr>
<td>Other Foils</td>
<td>0.67</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Discussion

Study 3 seems to support the hypothesis that the more numerate, or at least the more intelligent, are more likely to make conversions in a scenario that is important in the attribute framing literature. However, for this scenario, making the conversions did not appear to help them avoid the framing effect. In fact, conversions had no impact on the framing effect, but the more numerate, regardless of the fact that they may be making conversions, also happen to be more affected by the frame of the scenario. It is important to note that this study had a relatively large sample size (about 3 and a half times larger than Peters et al. 2006), and hence it appears that the differences between the results of the two studies are due to the scenarios, or the percentages in the scenarios, rather than due to chance.
General Discussion

The results of all 3 studies are summarized in Appendix D. The present paper, as well as many others could benefit from a better measure of numeracy and a better measure of conversions, perhaps resulting in fewer inconsistencies between studies.

The mechanisms described in the introduction of this paper appear to describe real behavior in some instances. That is, the more numerate do appear to make more conversions, leading them to have a smaller framing effect when the framing effect is measured within-subjects. The story is substantially more complicated between-subjects.

The within-subject results show that people are able to recruit mathematical tools, when present, in order to make consistent judgments. In fact, the recruitment of these tools for the purposes of consistency, rather than information integration, may result in equally biased judgments compared to when the tools are not recruited at all.

Consider a customer walking through a supermarket. He sees a pack of ground beef labeled 20% fat. Assume that if he were shown both the fat and the lean content, he would be at chance level to buy it (50%). However, shown the fat content, he thinks about the fat and grease and decides that the proportion of it is too high, and he decides he probably will not buy it but will come back later (let’s say the probability he will buy it is 20%). In the next aisle, he sees an identical package of beef labeled 80% lean. If he is less numerate, he may not realize the two packages of beef are the same, and may be more likely to purchase this apparently better package of beef (let’s say 80% likely).

If he is more numerate, he may realize the two packages are the same. Once he realizes this, he may think in order to maximize the decision: “I thought that last beef was
pretty fatty and was unlikely to buy it, and this beef has the same fat content, but now that it’s labeled differently I’m thinking it’s not so bad. Taking both thoughts to bear on the decision, maybe I’ll buy it after all.” The probability to buy here is greater than when he only saw the beef labeled fat (the probability to buy there was 20%), but less than when he just saw the beef labeled lean (the probability to buy there was 80%), so let’s say 50%. On the other hand, he may think in order to maximize consistency: “I thought that last beef was pretty fatty, and this beef has the same fat content, and even though it now seems like this beef is not so bad, I should act consistently, so I probably won’t buy it.” (Probability to buy the same as originally, at 20%).

In a subsequent experiment, it may be interesting to differentiate these two ways people use numeracy in the context of within-subject framing, perhaps resulting in a numeracy X desire for consistency interaction.

However, it appears that the relation between numeracy, conversions and between-subject framing is not a simple one. Even the previously documented subset of this relation: the relation between numeracy and the between-subject framing effect (Peters et al., 2006) is more complicated than expected from previous results.

One complication is the apparent differences in mechanisms between types of percentages. Numeracy moderates the effects of framing in a beneficial way when percentages are precise, as reported by Peters et al., 2006, and in another way when they are round, as found in the student scenario in Study 2 and the beef scenario in Study 3. However, looking at other scenarios in Study 2, this distinction does not exist. That is,
this interaction is totally absent for the snack scenario, and is sometimes in the opposite
direction for the drug scenario.

The tendency for the numerate to make conversions, too, is not consistent across
scenarios. While it is present both in the beef (Study 3) and the student scenarios (Study
2), it is absent when examining all 4 scenarios used in Study 2 in one model.

Attribute framing is usually studied in one context in each study, and often with
only one scenario (it is usually consumer behavior or medical decision making). The
underlying assumption is that the results of the study extend to all contexts, though
occasionally a cautionary note on generalization is added in the discussion. However, the
discrepancies described above appear to point to the idea that neither the mechanisms (in
this case, conversion) nor the empirical findings (in this case, the relation between
numeracy and framing) are necessarily consistent between different contexts. The
hypotheses posed in the introduction appear to describe behavior in the student framing
scenario used by Peters et al. (2006), and perhaps only when the percentage is precise,
but not when it is round. Similarly, the classic scenario about quality of beef appears to
produce larger framing effects in the more numerate when percentages are round, despite
the more numerates’ attempts to convert the information to the alternative frame. In order
to draw conclusions about attribute framing, it appears to be insufficient to examine only
one, or a small range of attribute framing scenarios because of this variation.

Levin et al (1998) created a distinction between the different framing effects,
making attribute framing one of the types of framing (risky choice and goal framing were
the other two). They justified this division by showing that there exists an independence
of effects across framing scenarios, i.e. the fact that a person is susceptible to risky choice framing does not necessarily mean s/he is susceptible to attribute framing. However, based on the results of Studies 2 and 3, it appears that even within attribute framing, not all scenarios behave in the same way. If more numerate people are less susceptible in some scenarios and more susceptible in others, by Levin and colleagues’ logic, this means that there exists an independence of effects within attribute framing. What exactly it is about the scenarios that causes the different behaviors is unclear, and remains to be studied in further work. However, it is clear that when studying the framing effect, one either has to take a large enough sample from the population of framing scenarios, or one has to limit the conclusions of each attribute framing study to the framing scenarios studied.

It appears that a prerequisite to answering this question is a study of framing effects across stimuli. In order to approach this topic from an exploratory direction, it may be necessary to find a large pool of attribute framing scenarios. It appears that there is no such pool in the literature, but a number of stimuli can be borrowed from previous experiments, and others may be created in order to supplement this pool. After getting consistent measures of framing effects, and numeracy X framing interactions by showing these framing stimuli to a range of subjects, a discrete version of a principal components analysis could be conducted.

Additive clustering is one procedure that is intended to solve such a problem (Shepard & Arabie, 1979). Additive clustering would take as input a similarity measure for the stimuli. Since here we are interested in the framing effect, one measure of
similarity could be the difference between framing effects of the stimuli. Without requiring any grouping variables a priori, this procedure would group the stimuli into possibly overlapping sets based on features that appear to drive the similarity structure. Therefore, when designing the stimuli, we may want to vary a large number of possible features, and by seeing which features the procedure returns we can arrive at a conclusion showing which features are relevant to the effect of interest. For example, features like domain (consumer vs medical vs environmental, etc), percentage(round vs precise, small vs large), and level of detail in the scenario may be manipulated on purpose, while others like length of stimulus may be incidental. The strength of the procedure is that, unlike a traditional design, it may return features that were not hypothesized to be important a priori. Since theoretical accounts of the framing effect generally assume that the framing effect is the same across stimuli, this is a great asset.

While designing a set of stimuli is an expedient approach, it may lack ecological validity. The results of an additive clustering procedure would be more meaningful if the input were responses to a naturally occurring sample of framing problems. While the literature does provide a sample of framing problems, there is no reason to believe that it is a good one. Researchers often pick stimuli that are relevant to what they are studying, or stimuli that are convenient. Fortunately, almost any percentage information can be turned into an attribute framing stimulus. A natural sample of attribute framing scenarios can therefore be derived by examining print media. This population of stimuli is a good one because it has ecological validity, and uses percentage information. A number of popular newspapers are currently available online. A computer program can be written to
scan online passages for percentage information. A participant pool can be used to find a relevant question to ask about this information. A pool of attribute framing scenarios can thereby be found that represents information conveyed to people on a regular basis.
References


Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational behavior and human decision processes, 76*(2), 149-188.


Footnotes

1 This type of model was introduced recently in psychology (Baayen et al., 2008). Specifically, it was introduced in the journal Memory and Language, which mainly concentrates on studies related to linguistics. As a result, it has chiefly been used to test for experimental effects while controlling for the effects of participants and words. However, it has been suggested that it should be used to control for effects of stimuli other than words in social psychology (Judd et al., 2012).

Imagine testing how quickly lexical decisions can be made. In such an experiment, a number of words and non-words are presented to a number of participants who decide as quickly as possible which are which. The response times will be slower for some people than for others, but they will also be slower for some words than others across participants. In order to make a good statistical test, one needs to account for these correlations, but a regular mixed effects model does not account for such effects.

In judgment and decision making research, we almost always use multiple participants, but we also often use multiple stimuli. For this reason, a crossed mixed effects model would seem to be beneficial because it can account for sources of covariation both within responses to stimuli and to participants. Indeed, several studies in judgment and decision making have used this model in analysis (Oppenheimer & Monin, 2009; Rudski et al, 2011; Baron & Ritov, 2009).

2 Since the design crossed participants with framing scenarios, we employed the crossed random effects model, but neither the intercept of judgments nor the slope of frame on judgments appeared to differ significantly across participants despite the large...
sample size (this was determined by fitting a crossed random effects model to judgments and applying an MCMC algorithm to find the 95% confidence interval for the variance accounted for by the subject factor, which included 0). Therefore, the model was reduced to a mixed model with intercepts and slopes for stimuli only.

Given the coding scheme described in Table 3, a 1.00 coefficient of frame indicates that changing the frame from negative to positive increased the judgment by 1 point on average. Increasing the percentage by one increased the judgment by 0.05 of a point on average. By dividing 1 by 0.05 we get 20, which is how many percent a change from negative to positive frame is equivalent to.
Appendix A: 17 Numeracy Items Used

1. Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?
   Answer: Half the time, 50%, any number between 490-510, 1:2

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?
   Answer: ___10___people

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?
   Answer: ___.1__%

4. Which of the following numbers represents the biggest risk of getting a disease?
   Answer: ___ 1 in 100 ___ 1 in 1000 _X_ 1 in 10

5. Which of the following numbers represents the biggest risk of getting a disease?
   Answer: ___ 1% _X_ 10% ___5%

6. If Person A's risk of getting a disease is 1% in ten years, and Person B's risk is double that of A's, what is B's risk?
   Answer: _2_% in_10_ years or 1% in 5 years

7. If Person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A, what is B's risk?
   Answer:  2 in 100 in 10 years or 1 in 50 in 10 years or 1 in 100 in 5 years.

8. If the chance of getting a disease is 10%, how many people would be expected to get the disease:
   A: Out of 100
   Answer: ____10__people
   B: Out of 1000
   Answer: __100___people

9. If the chance of getting a disease is 20 out of 100, this would be the same as having a __20__% chance of getting the disease.
10. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?
Answer: __5__ people

11. Which of the following numbers represents the biggest risk of getting a disease?
__X__ 1 chance in 12
_____ 1 chance in 37

12. Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not have a tumor. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. The table below summarizes all of this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

<table>
<thead>
<tr>
<th></th>
<th>Tested Positive</th>
<th>Tested Negative</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actually has a tumor</td>
<td>9</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Does not have a tumor</td>
<td>9</td>
<td>81</td>
<td>90</td>
</tr>
<tr>
<td>Totals</td>
<td>18</td>
<td>82</td>
<td>100</td>
</tr>
</tbody>
</table>

Answer: __9__ out of __18__

13. Imagine that you are taking a class and your chances of being asked a question in class are 1% during the first week of class and double each week thereafter (i.e., you would have a 2% chance in Week 2, a 4% chance in Week 3, an 8% chance in Week 4). What is the probability that you will be asked a question in class during Week 7?
Answer: __64__%

14. Suppose that 1 out of every 10,000 doctors in a certain region is infected with the SARS virus; in the same region, 20 out of every 100 people in a particular at-risk population also are infected with the virus. A test for the virus gives a positive result in 99% of those who are infected and in 1% of those who are not infected. A randomly selected doctor and a randomly selected person in the at-risk population in this region both test positive for the disease. Who is more likely to actually have the disease?
_____ They both tested positive for SARS and therefore are equally likely to have the disease
_____ They both tested positive for SARS, and the doctor is more likely to have the disease
They both tested positive for SARS and the person in the at-risk population is more likely to have the disease.

15. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
Answer: __5__ cents

16. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
Answer: __5__ minutes

17. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
Answer: __47__ days
Appendix B: Four Framing Scenarios Used In Study 1

Note: Questions are presented in the positive frame, with modifications to negative frame in parentheses.

1. Clearex is an acne cream available by prescription only in 12.8 ounce bottles. In clinical trials it caused no (mild and unpleasant) side effects in 86% (14%) of patients.

Please rate how safe this medication is compared to other medications of its type:

-3 Very Unsafe
-2
-1
0 Neither Unsafe nor Safe
+1
+2
+3 Very Safe

2. Dairy Farms has been in business since 1980. It has enjoyed moderate success in the last few years selling 95% fat free milk.

Please rate the nutritional value of this milk product:

-3 Very Low Nutritional Value
-2
-1
0 Neither Low nor High Nutritional Value
+1
+2
+3 Very High Nutritional Value
3. Health Bars can be bought in bulk if at least 131 of them are purchased. Each bar is a combination of nuts and dark chocolate, made with pure cane sugar. It is 75% sugar-free.

Please rate how healthy or unhealthy this snack is compared to other snacks of its type:

-3 Very Unhealthy
-2
-1
0 Neither Unhealthy nor Healthy
+1
+2
+3 Very Healthy

4. In an Introductory Psychology course with 245 students, John took a midterm that covered the history of psychological thought in the first half of the last century. John answered 87% of the questions correctly.

Please rate this student's performance:

-3 Very Low Performance
-2
-1
0 Neither Low nor High Performance
+1
+2
+3 Very High Performance
### Appendix C: Four Orders of Scenarios Used in Study 1

<table>
<thead>
<tr>
<th>Order 1</th>
<th>Order 2</th>
<th>Order 3</th>
<th>Order 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug: Clearex</td>
<td>Snack: Special Os</td>
<td>Student: John</td>
<td>Snack: Health Bars</td>
</tr>
<tr>
<td>Student: Peter</td>
<td>Student: Peter</td>
<td>Student: Emily</td>
<td>Student: Peter</td>
</tr>
<tr>
<td>Milk: White Valley</td>
<td>Snack: Tiny Bites</td>
<td>Milk: Dairy Farms</td>
<td>Milk: Moo Milk</td>
</tr>
<tr>
<td>Snack: Health Bars</td>
<td>Student: John</td>
<td>Drug: Miranol</td>
<td>Student: John</td>
</tr>
<tr>
<td>Student: Emily</td>
<td>Milk: White Valley</td>
<td>Drug: Geodane</td>
<td>Snack: Tiny Bites</td>
</tr>
<tr>
<td>Drug: Geodane</td>
<td>Student: Emily</td>
<td>Snack: Health Bars</td>
<td>Drug: Miranol</td>
</tr>
<tr>
<td>Drug: Miranol</td>
<td>Drug: Geodane</td>
<td>Snack: Special Os</td>
<td>Snack: Special Os</td>
</tr>
<tr>
<td>Milk: Dairy Farms</td>
<td>Drug: Miranol</td>
<td>Snack: Tiny Bites</td>
<td>Student: John</td>
</tr>
<tr>
<td>Student: John</td>
<td>Drug: Clearex</td>
<td>Milk: Moo Milk</td>
<td>Milk: Dairy Farms</td>
</tr>
<tr>
<td>Milk: Moo Milk</td>
<td>Milk: Dairy Farms</td>
<td>Milk: White Valley</td>
<td>Drug: Geodane</td>
</tr>
<tr>
<td>Snack: Tiny Bites</td>
<td>Milk: Moo Milk</td>
<td>Drug: Clearex</td>
<td>Drug: Clearex</td>
</tr>
<tr>
<td>Snack: Special Os</td>
<td>Snack: Health Bars</td>
<td>Student: Peter</td>
<td>Milk: White Valley</td>
</tr>
</tbody>
</table>

Table 13. Four randomly generated order conditions in experiment 1.

*Note.* Each participant viewed the scenarios in this order twice, either viewing the positive or the negative frame first.
Appendix D: Summarized Results of All Four Studies
<table>
<thead>
<tr>
<th>Study</th>
<th>Hyp 1</th>
<th>Hyp 2</th>
<th>Hyp 3</th>
<th>Hyp 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>The more numerate had smaller framing effects</td>
<td>Errors on the foil of interest in the second frame predicted frame consistency</td>
<td>The more numerate had more errors on the foils of interest in the second frame</td>
<td>Errors on the foil of interest mediated the relation between numeracy and the framing effect</td>
</tr>
<tr>
<td>Study 1</td>
<td>(within-subj)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1</td>
<td>The more numerate had smaller framing effects</td>
<td>Errors on the foil of interest in the first frame predicted the framing effect (p = 0.07)</td>
<td>The more numerate had more errors on the foil of interest in the first frame</td>
<td>Not attempted because no effects reached significance</td>
</tr>
<tr>
<td>Study 1</td>
<td>(between-subj; round and precise)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1</td>
<td>The more numerate had smaller framing effects</td>
<td>Cannot estimate because no errors on student scenario in the positive frame</td>
<td>The more numerate had more errors on the foil of interest in the first frame</td>
<td>Cannot estimate as in Hyp. 2.</td>
</tr>
<tr>
<td>Study 1</td>
<td>(between-, student scenario)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>The numerate had smaller framing effects (p = 0.15)</td>
<td>Recall of alternative frame predicted the framing effect</td>
<td>The more numerate were more likely to recall the alternative frame</td>
<td>Not attempted because two effects did not reach significance</td>
</tr>
<tr>
<td>Study 2</td>
<td>(between-, all scenarios)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>The more numerate had smaller framing effects (significant precise only)</td>
<td>Cannot estimate because no errors were made on the student scenario in the positive frame</td>
<td>The more numerate were more likely to recall the alternative frame</td>
<td>Cannot estimate because no errors were made on the student scenario in the positive frame</td>
</tr>
<tr>
<td>Study 3</td>
<td>The more numerate had smaller framing effects (effect in the opposite direction)</td>
<td>Errors on the foil of interest predicted the framing effect</td>
<td>The more numerate had more errors on the foil of interest in the first frame (p = 0.17 when controlling for intelligence)</td>
<td>Not attempted because two effects did not reach significance</td>
</tr>
<tr>
<td>Study 3</td>
<td>(between-, beef scenario; round %)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 14. Summarized results of Studies 1-3.
Note. Bold black font on white means there was full support for the hypothesis, even when controlling for the appropriate proxies of intelligence ($|t|>2$). A regular font in block indicates directional results. A light gray font means the effect is absent ($|t|<1$). A gray background indicates the analysis was impossible.
Appendix E: Correlations and Scatterplots of Individual Difference Variables In Studies 1 and 2
Figure 4. Correlations and Scatterplots of Individual Difference Variables in Study 1.

Note. By taking the variable numbers, and finding the corresponding position in the matrix, one can find the correlation, or the scatterplot. The variables in order are: numeracy, working memory, vocabulary, Verbal SAT, Math SAT, and ACT scores. For
example to find the correlation and scatterplot between variable 1 (numeracy) and variable 3 (Math SAT), I would go to column 1, row 3, and row 3, column 1.
Figure 5. Correlations and Scatterplots of Individual Difference Variables in Study 1.

Note. By taking the variable numbers, and finding the corresponding position in the matrix, one can find the correlation, or the scatterplot. The variables in order are: Raven’s progressive matrices score, numeracy, Math SAT, Verbal SAT, and ACT scores. For
example to find the correlation and scatterplot between variable 1 (Raven’s matrices) and variable 3 (Math SAT), I would go to column 1, row 3, and row 3, column 1.