

Subjective integration of probabilistic information from experience and description

Yaron Shlomi*

Abstract

I report a new judgment task designed to investigate the subjective weights allotted to experience and description when integrating information from the two sources. Subjects estimated the percentage of red balls in a bag containing red and blue balls based on two samples from the bag. They experienced one sample by observing a sequence of draws and received a description of the other sample in terms of summary statistics.

The results of two experiments show that judgments were more sensitive to the experienced sample compared to the described one for most subjects, although others showed the opposite bias. The bias toward experience varied as a function of the presentation order of the two samples in Experiment 1 and the presentation format of the description in Experiment 2.

The integration of description and experience exemplifies tasks that require integration of information obtained from different sources and in different formats. Informed by the findings reported in this study, I identify some directions for future research on human information integration.

Keywords: description vs. experience, information integration, subjective probability judgment, numerical presentation format

1 Introduction

Human judgment and decision making can be guided by two distinct sources of information, personal experience or description. Experience refers to observing information directly whereas description refers to information that has been observed and abstracted by a source other than the judge/decision maker. To exemplify the distinction, consider the information guiding physicians: they obtain experience from interacting with patients whereas they obtain description by reading professional literature.

Intuition suggests that humans integrate information from the two sources in forming judgments and making decisions (e.g., physicians integrate what they have learned from description and experience in choosing a treatment plan). Although such integration informs potentially consequential decisions, very little is known about

its quality and the psychological processes that underlie it. The purpose of the current research was to assess the influence (i.e., relative weights) that people allot to description versus experience. The research focus relates to the description-experience gap (Barron, Leider & Stack, 2008; Hertwig, Barron, Erev & Weber, 2004). The theoretical framework (i.e., the notion of weighting) relates to information integration in other contexts including subjective averaging (Anderson, 1968; Levin, 1975), belief updating (Wallsten, 1972) and using advice (Yaniv & Kleinberg, 2000).

The paper is organized as follows. Research pertinent to integrating description and experience is reviewed in the next section. I then report two experiments. In the general discussion I summarize and interpret the findings and discuss their implications for research on processing description and experience.

1.1 Previous research on integrating description and experience

Following previous literature, experience and description are defined as different methods of obtaining information about a population of outcomes (e.g., Barron et al., 2008; Hau, Pleskac, Kiefer, & Hertwig, 2008). Experience refers to information obtained from sampling individual outcomes from the population. Description refers to information obtained from a numerical summary of a sample (i.e., 80% of the chips are red).

This article is based on a dissertation submitted in partial fulfillment of the doctoral requirements at the University of Maryland–College Park. Portions of this research were presented as a poster at the 2009 annual meeting of the Society for Judgment and Decision Research, Boston, Massachusetts.

I extend my gratitude to Thomas S. Wallsten (chair), Tom Carlson, Michael Dougherty, Rebecca Hamilton, and Cheri Ostroff for their service on my dissertation committee. I also thank Joshua Boker, Ezra Geis, Leda Kaveh, Marissa Lewis, Stephanie Odenheimer, Lauren Spicer, Herschel Lisette Sy, and Kimberly White for their help in collecting the data.

Copyright: © 2014. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Department of Industrial Engineering and Management, Shenkar College of Engineering and Design, Ramat Gan, Israel. Email: yshlomi@shenkar.ac.il.

The literature review is aimed at motivating four questions about the subjective weighting allotted to experience and description. (1) Do people allot equal weights to description and experience, or do they systematically allot more weight to one of these sources? (2) Does the allocation of the weights depend on the particular outcome associated with each of the information sources (i.e., the information assignment)? (3) Does the subjective allocation of the weights depend on whether experience is obtained before versus after description (i.e., the source presentation order)? (4) Do the weights depend on the presentation format of the description? The motivation for these questions is reviewed in the next few paragraphs.

1.1.1 Equal versus unequal weighting

Evidence obtained by Newell and Rakow (2007) suggests that people are more sensitive to experience than to description. They provided people with a description, and then tested whether they were affected by experience containing the same information. Specifically, they provided subjects a description of a die with four black faces and two white faces and then asked subjects to predict the outcome of rolling the die (i.e., whether it would land on a black or a white face). Subjects were told that the die was fair and unbiased. The crucial manipulation was whether or not subjects observed the outcome after each prediction; i.e., whether or not they received experience.

Newell and Rakow (2007) found that subjects who observed the outcomes made correct predictions (i.e., they predicted the black face of the die) more often than those who did not observe the outcomes. Clearly, subjects relied on experienced information to revise their predictions.

Newell and Rakow's (2007) results suggest that, when observers are presented with both description and experience, they integrate information from the two sources, and are more sensitive to experience than description. Experience, even when the information it provides is redundant (i.e., it is identical to the description) yields more correct predictions.

Evidence related to the distinction between description and experience motivates a more specific expectation that experience will be weighted more heavily than description in the integrated output. This evidence is borrowed from research on using advice (e.g., Yaniv & Kleinberg, 2000), frequency versus probability formats (e.g., Gigerenzer & Hoffrage, 1995) and product preferences given exposure to trial versus ads (e.g., Hamilton & Thompson, 2007).

1.1.2 Information assignment

The information assignment refers to the association of the two samples with the two sources. One assignment consists of a description, *A*, and an experience, *B*, where *A*

and *B* are probabilistic outcomes. Another assignment, obtained from reversing the first, consists of a description, *B*, and an experience, *A*. Do the two assignments yield similar responses?

Newell and Rakow (2007) did not counterbalance the assignment of the information units to the two sources. Subjects received a description of sample *A* and experienced sample *B*. They did not receive the reverse assignment. Thus, it is unknown whether the behavior observed by Newell and Rakow should be attributed to processing description versus experience or to processing particular values of *A* and *B*. The generalizability of evidence from other tasks involving information integration is unclear.

1.1.3 Source presentation order

Subjective integration is sensitive to the sequence of presenting information from the two sources. In a relevant study, Barron, Leider, and Stack (2008; Experiments 1, 2, and 3) asked subjects to make 100 risky choices between two outcome distributions. Subjects received outcome information (experience) that was contingent on each choice. In addition to experiencing the outcome distributions, subjects read a warning about a large but unlikely loss associated with one of the distributions (i.e., subjects received a description). Crucially, one group of subjects received the warning before making the first choice, and a second group received it after making the 50th. Although the two groups had the same information after the 50th trial (i.e., choice), their choices in trials 51–100 were not comparable. Specifically, subjects who were warned before making the first choice were choosing the risky option approximately 25% more often than those who were warned after the 50th.

Additional evidence reported by Barron and colleagues (2008; Experiment 4) indicates that the source presentation order depends on the operational definition of experience. In Experiments 1, 2, and 3, subjects made 100 choices between two outcome distributions. In contrast, in the first 50 trials of Experiment 4 subjects merely observed a sequence of monetary gains sampled from each of the two distributions. One subject group obtained description (i.e., the warning) followed by experience; a second subject group obtained the information in the reverse sequence. Then, in trials 51–100, after they had obtained information from both sources, subjects made a set of consequential choices. There was no evidence that choices were affected by the presentation sequence of the information from the two sources.

In sum, the evidence presented by Barron and colleagues (2008) suggests that the effect of the presentation sequence on the attention allotted to description and experience depends on their operational definition. Furthermore, whereas Barron and colleagues implemented a

choice task, I implement a judgment task. Thus, a clear prediction about the effect of the source presentation sequence cannot be formulated.

1.1.4 Presentation format of the description

Descriptions of uncertainty can be presented in various formats (e.g., percentages, relative frequencies, graphs). However, the role of the presentation format on the allocation of weights in integrating description with experience has not been investigated.

Previous research indicates that processing of uncertainty depends on the format used to present the uncertainty (e.g., Gottlieb, Weiss & Chapman, 2004; Johnson, Payne, & Bettman, 1988). Moreover, research on Bayesian inference has led to claims that information presented in a relative frequency format leads to more accurate judgments than that presented in percentage or probability formats (e.g., Gigerenzer & Hoffrage, 1995). Such claims could be interpreted to imply that people are more sensitive to description (i.e., they will allot more weight to it) if it is presented as a relative frequency compared to a percentage.

1.2 New methodology for investigating the integration of description and experience

I constructed a judgment task that required subjects to integrate information from two samples to yield an estimate of the corresponding population average. The two samples provided information about the composition of a bag of red and blue chips (for a related design, see Phillips & Edwards, 1966; Pitz, Dowling, & Reinhold, 1967). In any one trial, subjects experienced a sample by observing a sequence of sampled chips and received a description of another sample in the form of a summary of its composition. After receiving the information in both samples, subjects estimated the percentage of red chips in the bag (i.e., the population parameter). Subjects were told that the description was trustworthy, and that the two samples associated with each bag consisted of the same number of chips.

The task was designed such that each pair of samples from a particular bag appeared in two experimental conditions over the course of a session. In one condition, sample *A* was experienced and sample *B* was described and in the second condition, sample *A* was described and sample *B* was experienced.

1.3 Overview of experiments and hypotheses

The judgment task was implemented in two experiments. The information assignment was manipulated in both experiments. The presentation order of the two sources was manipulated in Experiment 1, and the format of the description was manipulated in Experiment 2.

I summarize the predictions for the experiments as follows. (1) Experience will be weighted more heavily than description. (2) It is unclear whether the weight allotted to a particular information source depends on the outcome assigned to that source (i.e., the information assignment). (3) The effect of the presentation sequence on the weights is unclear. (4) People will allot more weight to description (and less weight to experience) if it is presented as a relative frequency compared to a percentage.

2 Experiment 1

The purpose of the experiment was to assess the weights allotted to probabilistic information obtained from description and from experience. I examined whether the integrator's use of the information depended on the source that provided it (i.e., the information assignment), and whether the integration was sensitive to the presentation order of the two sources.

2.1 Method

2.1.1 Subjects

One hundred sixty-two University of Maryland, College Park undergraduate students participated for course credit. In addition, they received a reward contingent on the accuracy of their judgment (see below).

2.1.2 Stimuli

Two sets of bags were used in the experiment. Bags in the "identical-percentage" set were associated with pairs of samples that contained an identical percentage of red chips. Bags in the "different-percentage" set were associated with pairs of samples that differed in the percentage of red chips. There were 10 and 18 bags in the identical- and different-percentage sets, respectively. All of the identical-percentage bags and 14 different-percentage bags were used in the experiment. The remaining four different-percentage bags were used for practice. The sample size (i.e., the number of chips in the samples) ranged over trials from 8 to 13 and was always the same for a pair of samples in a given trial. The sample and population percentages of red chips ranged from 14% to 86% (see Appendix A).

2.1.3 Procedure

Subjects were presented with instruction screens. The relevant instruction screens are displayed in Appendix B. The instructions were followed by one trial with each of the practice bags. Subjects typed their responses, and were then prompted to ask the experimenter for clarifications about the task. The responses obtained on the practice trials were excluded from the data analyses. After this, subjects completed two trials with each of the experimental bags for a total of 52 trials (i.e., $4 + 2 \cdot 24$).

The practice trials and experimental trials were identical in design. Subjects initiated each trial by clicking a button. Each trial consisted of three parts. First, subjects clicked a button to draw one chip from one sample, and continued clicking the button until they had viewed each of the chips in that sample. The chip appeared on the display 500 ms after each click, and remained visible until the next button click. Second, subjects clicked a button to receive the description of the second sample. The description consisted of a picture of “Mr. Rick” (i.e., the source of the description), the number of chips that he sampled, and the percentage of red chips he observed. This information remained visible until the subject’s next click. Third, subjects typed their estimate of the percentage of red chips in the bag. The experiment was programmed so that only integers in the [1, 99] interval were accepted.

Subjects were randomly allocated to two experimental conditions. One group of subjects ($n = 82$) obtained information from experience and then from description (i.e., Experience-1st). A second group ($n = 80$) obtained information on each trial from description and then experience (i.e., Experience-2nd). Two bag presentation sequences were counterbalanced across subjects. The sequences were arranged so that the first and second presentation of each bag occurred in the first and second block of 24 consecutive experimental trials, respectively. The presentation of bags in the same- and different-percentage sets was intermixed within each block.¹

The two blocks differed from each other in the information assignment of each of the different-percentage bags. In one block, the extreme sample associated with each bag was experienced and the moderate sample associated with each bag was described (e.g., $P_E = 80\%$ and $P_D = 60\%$, where P_E and P_D correspond to the experienced and described samples). In the second block, this assignment was reversed. The order of the two blocks was counterbalanced across subjects.

The computer scored the accuracy of the subject’s re-

sponse on each trial using the following rule, $s = 100[1 - (R - R^*)^2]$, where R and R^* correspond to the observed response and the mean of the two samples, respectively. At the end of the subject’s session, the computer computed the subject’s average score from the scores associated with the 48 experimental trials. The average score is bounded in [0, 100]; the value of s determined the probability that the subject earned a reward (i.e., a commuter’s mug). Since subjects did not receive any feedback in the course of the experiment, the reward could not affect the data analyses and is not considered further.

2.2 Results

I computed the weight allotted to experience (w) for each item using $R = wE + (1 - w)D$, where R is the subject’s response, and E and D are the experienced and described percentages (for a similar computation, see Soll & Mannes, 2011). The terms are rearranged to obtain $w = (R - D)/(E - D)$. Responding with the described percentage, the average, or the experienced proportion yields $w = 0, .5$, and 1, respectively.

The weights indicated more sensitivity to experience than description. The average weight ($M = .60$, $SE = .02$) was significantly greater than .5, $t(161) = 5.91$, $p < .001$. Greater sensitivity to experience was observed in the Experience-2nd sequence ($M = .63$, $SE = .02$) compared to the Experience-1st sequence ($M = .56$, $SE = .02$). Deviations toward experience were greater when the experienced sample was moderate ($M = .61$, $SE = .02$) than when it was extreme ($M = .59$, $SE = .02$). The presentation sequence yielded a significant effect, $F(1,160) = 4.74$, $p < .05$. The effect of assignment (of extreme/moderate to experience/description) was almost significant, $F(1, 160) = 3.65$, $p = .06$, and the interaction of assignment with the presentation sequence was not significant, $F < 1$.

Inspection of the data indicated that on average, only 79% (i.e., 22 out of 28) of the subjects’ judgments were bracketed by the two sample proportions. Stated differently, 21% of the judgments were outside the interval bounded by the sample proportions. This finding is difficult to interpret in the context of the assumption that subjects allocate their attention between the two sources.

In a follow up analysis I focused only on the responses inside the bracket. The average weight of the responses within the bracket ($M = .55$, $SE = .01$) was significantly greater than .5, $t(161) = 4.56$, $p < .001$. Thus, the general pattern of the deviations toward experience is not merely the product of responding outside the bracket. The weights were not significantly affected by order, assignment, and the order by assignment interaction, all $F_s < 1$.

¹The same-percentage bags shown in Appendix A were included in the design for completeness but do not contribute to the analyses. Thus, they will not be mentioned further.

2.2.1 Individual differences²

The average weight of most of the subjects (66%) tended toward experience (i.e., $w > .5$). To examine whether the average weight of each subject was significantly different from .5, I computed, for each subject, a Wilcoxon test (used because of long tails on the distributions of weights) and its associated p value (one-tailed), and asked whether the number of significant results ($p < .05$) in each direction exceeded the number expected by chance.

Sixty-five out of 162 subjects had a statistically significant deviation toward experience. This is significantly more than the expected 8.1 (0.05·162). This is significantly more, approximately $p < .001$ by a one-tailed binomial test. A significant bias toward description was observed for 19 subjects. This result is also significant at $p < .001$. A similar pattern was observed within each presentation sequence.

The individual difference analysis yielded a statistically significant asymmetry for half of the subjects. The majority of these subjects give more weight to experience, but some give more weight to description. No reliable bias was found for the other half of the subjects.

2.3 Discussion

Experiment 1 assessed the allocation of weights in the subjective integration of description- and experience-based probabilistic outcomes.

The weights indicated more sensitivity to experience than description for most subjects. That the weights are sensitive to the presentation sequence (i.e., whether experience precedes description or follows it) is open to multiple interpretations. Presumably, the integration involves translating frequencies to percentages and/or percentages to frequencies. The presentation sequence might prime the judges' choices of a particular translation (for similar reasoning, see Hogarth & Einhorn, 1992).

3 Experiment 2

Experiment 2 served two purposes. One was to examine whether the allocation of the weights observed in Experiment 1 was the consequence of the presentation format of the description, namely, presentation as a percentage. Thus, I tested whether the subjective weights vary as a function of the percentage format (as used in Experiment 1) versus the relative frequency format. Claims about the role of format in Bayesian inference (e.g., Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) suggest the frequency format would lead to more balanced allocation

of weights than the percentage format. The second purpose of the experiment was to explore whether the allocation of the weights is related to the perceived trustworthiness of the source of the description.

3.1 Method

3.1.1 Subjects

One hundred and one undergraduate students participated in return for course credit.

3.1.2 Stimuli

The stimuli differed from those of Experiment 1 in four ways. (1) There were eight practice and sixteen experimental bags. (2) The two samples associated with each bag always consisted of unequal but categorically similar proportions (i.e., both samples had the same majority color). (3) The samples always consisted of 13 chips. (4) The sample and population percentages ranged from 0 to 100 and from 4% to 96%, respectively (see Appendix C).

3.1.3 Design

The description format, percentage or frequency, was manipulated between-subjects. The presentation sequence was not manipulated; only the Experience-2nd sequence (description then experience, the sequence yielding a greater weight of experience relative to description) was used. Other details of the design were similar to those of Experiment 1.

3.1.4 Procedure

After receiving instructions, subjects completed eight practice trials and 32 experimental trials (i.e., there were two replications of each experimental bag). The responses obtained on these practice trials were excluded from the data analyses and will not be mentioned further. After the practice trials, subjects were prompted to ask the experimenter for clarifications about the task, and continued to the experimental trials. The familiarization and experimental trials were identical in design.

Each trial consisted of three parts (as in Experiment 1). Subjects clicked a button to receive the description of one sample. The description consisted of the following text, "Mr. Rick sampled 13 chips. x of the chips in Mr. Rick's sample were red." Depending on the condition x was either a percentage (e.g., 62%) or number (e.g., 8). The description was displayed for 2500 ms.

Subjects clicked a button to obtain experience (i.e., to start drawing the chips) in the second sample. Each chip was displayed for 1000 ms, and the inter-chip-interval was 2000 ms.

²Analysis of the individual differences was requested by the Editor.

Subjects provided their estimates with a slider anchored “0% Red” on the left end, “50% Red” in the middle, and “100% Red” on the right end. Thus, unlike Experiment 1, subjects were not restricted to the [1, 99] interval and they were not required to type a numerical response.

After completing the last trial, subjects judged the following statement, “I trusted Mr. Rick to provide reliable information about the bag of chips.” Subjects responded by marking a 5-point scale labeled with “Completely disagree”, “Somewhat disagree”, “Neutral”, “Somewhat agree” and “Completely agree”.

3.2 Results

As in Experiment 1, the weights (w 's) indicated more sensitivity to experience than description. Specifically, the average weight ($M = .63$, $SE = .03$) was significantly greater than .5, $t(100) = 4.7$, $p < .001$. Subjects who were presented with description in the percentage format placed more weight on experience ($M = .70$, $SE = .04$) than those presented with the relative frequency format ($M = .56$, $SE = .04$). The weights when experience was moderate and extreme were .66 ($SE = .05$) and .60 ($SE = .06$) respectively. Format (percentage vs. frequency) yielded a significant effect on w , $F(1, 99) = 7.07$, $p < .01$. Assignment and the assignment by format interaction were not significant, $ps > .1$.

On average, only 64% (10 out of 16) of the responses were bracketed by the two sample proportions. This percentage is lower than the 79% observed in Experiment 1.

As in Experiment 1, the weights ($M = .55$, $SE = .01$) of the responses inside the bracket deviated toward experience, $t(100) = 4.2$, $p < .001$. Greater weights (i.e., in the direction of experience) were observed in the percentage format ($M = .60$, $SE = .02$) than in the relative frequency format ($M = .53$, $SE = .02$) and when experience was extreme ($M = .63$, $SE = .02$) versus moderate ($M = .49$, $SE = .02$). The effects of information assignment and format were significant, $F(1, 99) = 39.44$, $p < .001$, and $F(1, 99) = 6.45$, $p < .05$, respectively. The interaction of assignment and format was not significant, $F(1, 99) = 1.41$, $p = .24$.

3.2.1 Individual differences

Most (66 out of 101) subjects had an average weight that tended toward experience (i.e., $w > .5$). Following the procedure in Experiment 1, I found that 35 subjects (out of 101) had a statistically significant bias ($p < .05$ by a Wilcoxon test, as in Experiment 1) toward experience, which exceeds the expected number of 5.05 at $p < .001$ by a one-tailed binomial test. Nine subjects showed a bias toward description, which is not quite significantly greater than 5.05 ($p = .066$), but not much lower as a proportion

than the 19 cases in Experiment 1 (9% vs. 12%), which had a larger sample and used both sequences.

3.2.2 Perceived reliability of the description

After completing the integration task, subjects judged the assertion that Mr. Rick provided reliable information about the bag of chips. Subjects responded by rating whether they completely disagreed, somewhat disagreed, were neutral, somewhat agreed, or completely agreed with the statement. For analyses purpose, the subjects' ratings were coded on a scale from -2 (complete disagreement) through 0 (neutral) to $+2$ (complete agreement).

The trust ratings were related to the observed judgments. Specifically, the weights were negatively related to the trust ratings (Spearman's $r = -.50$, $p < .001$). In other words, subjects' reliance on experience was inversely related to their trust in Mr. Rick.

Mr. Rick was perceived as more trustworthy when he presented the sample outcome as a frequency rather than a percentage. The mean (median) trust ratings in the percentage and fraction formats were .4 (1.0) and 1.2 (1.0), respectively. The difference between the ratings in the two formats was significant by a Mann-Whitney U test, $p < .05$.

3.3 Discussion

The judgments again predominantly deviated in the direction of the experienced sample. The weights were sensitive to the description format. Ratings of Mr. Rick's trustworthiness were affected by the description format: his descriptions, conveyed in the frequency format were perceived as more reliable than the same descriptions in the percentage format. Ratings of Mr. Rick as more trustworthy were related to lower weight than ratings of him as less trustworthy.

The effect of the presentation format on both the weights and the trust ratings suggests that the frequency format is better than the percentage format for conveying description. Processing frequencies may be more similar, or perhaps identical, with processing experienced information compared to processing percentages (Gottlieb et al., 2007). Thus, the advantage of the frequency format is attributed to the similarity (compatibility) of the processes that operate on description and on experience. Processing compatibility, in turn, might be tied to processing fluency; if so, subjects may judge the trustworthiness of Mr. Rick by assessing how fluently they process information that he provides (Werth & Strack, 2003).

4 General discussion

This research investigated judgments informed by integrating description- and experience-based probabilistic outcomes. Judgments were biased in the direction of the experienced outcome. There was some evidence that the judgments were affected by the distribution of the information across the sources; i.e., whether they were informed by an extreme experience and moderate description or by the opposite assignment (Experiment 1). The weights varied as a function of the source presentation sequence (Experiment 1, more attention to experience when it came second) and the numerical format of the description (Experiment 2, more attention to description when it was presented as a frequency).

The weighted average operation is one conceptualization of the subjective aggregation of information from description and experience. Several subjective differences between the two sources might determine their processing weights. Such dimensions include precision (e.g., Du & Budescu, 2005), concreteness (Hamilton & Thompson, 2007), the effort/fidelity involved in coding information from different formats (e.g., Johnson, Payne, & Bettman, 1988), and credibility (Yaniv & Kleinberger, 2000). Presumably, the integration process allots more weight to the source associated with higher values on these dimensions and less weight to the source associated with lower values.

The judgments occasionally fell outside the interval bounded by the described and experienced outcomes, leading to weights outside the 0-1 interval. This finding indicates that the weighted average principle is probably inadequate to describe the process underlying these responses.

Subjects might assume that the two estimates might not always bracket the true proportion. Thus, responses outside the bracket might reflect attempt to guess when that might happen (Soll & Mannes, 2011). This guessing strategy is distinct from the weighted average strategy.

The average response of most of the subjects indicates more reliance on experience compared to description. A more rigorous analysis examined the statistical reliability of the asymmetry in the subjects' weights. More subjects were associated with a reliable bias toward experience than toward description, although some were biased toward description. In addition, a sizable group of subjects showed no evidence of a reliable asymmetry.

Future research is necessary to elucidate which strategies subjects use to represent and weight the information, and how strategy choice is affected by the task-related variables (e.g., presentation sequence). Identifying the factors that subjects use to assign the weights is also important.

References

- Anderson, N. H. (1968). Averaging of space and number stimuli with simultaneous presentation. *Journal of Experimental Psychology*, *77*, 383–392.
- Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience of a warning's impact: sex, drugs and rock-n-roll. *Organizational Behavior and Human Decision Processes*, *106*, 125–142.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.
- Du, N., & Budescu, D. V. (2005). The effects of imprecise probabilities and outcomes in evaluating investment options. *Management Science*, *51*, 1791–1803.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.
- Gottlieb, D. A., Weiss, T., & Chapman, G. B. (2007). The format in which uncertainty is presented affects decision biases. *Psychological Science*, *18*, 240–246.
- Hamilton, R. W., Thompson, D. V. (2007). Is there a direct substitute for direct experience? Comparing consumer's preferences after direct and indirect product experience. *Journal of Consumer Research*, *34*, 546–555.
- Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience outcome in risky choice: the role of sample size and experience probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518.
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effects of rare events in risky choice. *Psychological Science*, *15*, 534–539.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.
- Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, *42*, 1–21.
- Levin, I. P. (1975). Information integration in numerical judgments and decision processes. *Journal of Experimental Psychology: General*, *104*, 39–53.
- Newell, B. R. & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin and Review*, *14*, 1133–1139.
- Philips, L. D., Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.
- Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*, 81–102.

Wallsten, T. S. (1972). Conjoint-measurement framework for the study of probabilistic information processing. *Psychological Review*, 79, 245–260.

Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory*, 4, 411–419.

Yaniv, I. & Kleinberger, E. (2000). Advice taking in decision making. Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, 83, 260–281.

Appendix A: Stimuli in Experiment 1

Bag type	Bag	Sample n	Sample % red chips		Benchmark ^a
			Sample 1	Sample 2	
Experimental trials					
Identical ^b	1	7	14	14	14
	2	7	43	43	43
	3	7	57	57	57
	4	7	86	86	86
	5	9	11	11	11
	6	9	89	89	89
	7	11	9	9	9
	8	11	91	91	91
	9	13	8	8	8
	10	13	92	92	92
Different ^c	11	7	14	57	36
	12	7	86	43	65
	13	9	22	33	28
	14	9	78	67	73
	15	9	33	56	45
	16	9	67	44	56
	17	11	18	36	27
	18	11	82	64	73
	19	11	27	64	46
	20	11	73	36	55
	21	13	15	46	31
	22	13	85	54	70
	23	13	31	62	47
	24	13	69	38	54
Practice trials					
Different ^c	P1	6	17	33	25
	P2	6	83	50	67
	P3	14	21	64	43
	P4	14	79	43	61

^a Expected estimate of red chips in the bag (%).

^b Identical-percentage bags.

^c Different-percentage bags.

Appendix B: General instructions

Welcome to this experiment.

The purpose of this experiment is to investigate how people combine information from two sources.

In the experiment, you will be presented with a set of bags that contain red and blue chips. You will receive information about each bag from *two* sources. Your task is to estimate the proportion of red chips in each bag.

In the following screens, you will receive specific information about the experiment. The two sources of information about each bag will be introduced to you, and you will receive specific instructions on your task.

Instructions about experience

Sources of information about each bag of chips

You will obtain some information about each bag by drawing a number of chips from it. The computer will determine how many chips you will draw from each bag.

You will click the button to draw one chip. After observing whether the chip is red, you will need to click the button to draw another chip. The computer will return the chip to the bag after each draw.

You will repeat the process of clicking to draw a chip and observing the chip's color a number of times. Again, the number of chips you will observe is determined by the computer.

Instructions about description

Sources of information about each bag of chips

Mr. Rick will also provide you with information about each bag of chips.

He drew chips from each bag. He will tell you how many chips he drew from each bag, and the proportion of red chips that he saw.

You can assume that Mr. Rick provides reliable information.

Additional instructions

The computer will decide whether you draw the chips first and then receive Mr. Rick's information, or whether you receive Mr. Rick's information first and then draw the chips.

You will be inspecting many bags in this experiment. All of the bags contain 500 chips, but they differ in the proportion of red and blue chips. You should assume that the bags contain only red and blue chips, and that each bag has at least one red chip and at least one blue chip.

Appendix C: Stimuli in Experiment 2

Bag	Sample % red chips		Benchmark estimate ^a (%)
	Sample 1	Sample 2	
Experimental trials			
1	0	23	12
2	100	77	88
3	8	15	12
4	92	85	88
5	0	38	19
6	100	62	81
7	8	31	19
8	92	69	81
9	15	38	27
10	85	62	73
11	8	46	27
12	92	54	73
13	23	46	35
14	77	54	65
15	31	38	35
16	69	62	65
Practice trials			
P1	0	8	4
P2	100	92	96
P3	23	15	19
P4	77	85	81
P5	23	31	27
P6	77	69	73
P7	38	46	42
P8	62	54	58

^a Expected estimate of red chips in the bag (%). Note. Only different-percentage bags were used in this experiment. For all of the bags, n=13.