

A Case-Based Model of Probability and Pricing Judgments: Biases in Buying and Selling Uncertainty

Lyle A. Brenner

Warrington College of Business Administration, University of Florida, Gainesville, Florida 32611, lbrenner@ufl.edu

Dale W. Griffin

Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada,
dale.griffin@sauder.ubc.ca

Derek J. Koehler

Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, dkoehler@waterloo.ca

We integrate a case-based model of probability judgment with prospect theory to explore asset pricing under uncertainty. Research within the “heuristics and biases” tradition suggests that probability judgments respond primarily to case-specific evidence and disregard aggregate characteristics of the class to which the case belongs, resulting in predictable biases. The dual-system framework presented here distinguishes heuristic assessments of value and evidence strength from deliberative assessments that incorporate prior odds and likelihood ratios following Bayes’ rule. Hypotheses are derived regarding the relative sensitivity of judged probabilities, buying prices, and selling prices to case- versus class-based evidence. We test these hypotheses using a simulated stock market in which participants can learn from experience and have incentives for accuracy. Valuation of uncertain assets is found to be largely case based even in this economic setting; however, consistent with the framework’s predictions, distinct patterns of miscalibration are found for buying prices, selling prices, and probability judgments.

Key words: probability; finance; asset pricing; decision analysis; prospect theory; value function; forecasting

History: Received July 15, 2010; accepted May 31, 2011, by Brad Barber, Teck Ho, and Terrance Odean, special issue editors. Published online in *Articles in Advance* October 28, 2011.

Introduction

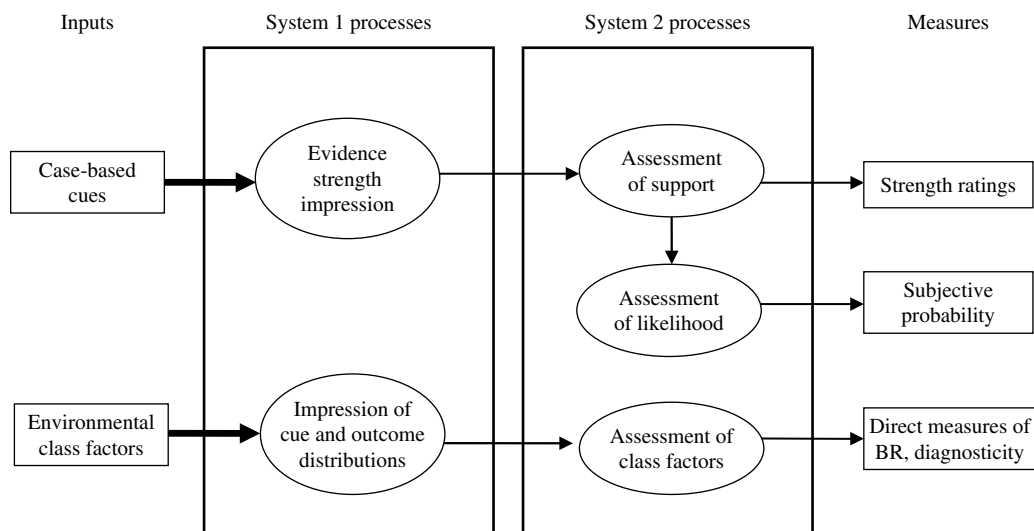
Judgments of probability offered by both experts and laypeople exhibit systematic biases, whether those biases are measured in terms of departures from normative standards or from actual outcomes (Tversky and Kahneman 1974, Griffin and Tversky 1992, Koehler et al. 2002). These biases in probability judgments have been taken as signature evidence for the operation of heuristic assessment processes. Much influential research in behavioral finance has attempted to tie these probability judgment biases to anomalies observed in the financial markets (e.g., Barberis et al. 1998, Daniel et al. 2001, Odean 1999). However, the question of whether biases found in judgments of probability are moderated in economic valuation tasks, such as the subjective evaluation of uncertain assets (asset pricing), has received little direct attention. Most related research has focused on whether biases found in probability judgment also exist—in a broad qualitative sense—in economic valuation tasks (e.g., Bloomfield 1996, Camerer 1992,

Fox et al. 1996, Fox and Tversky 1998, Kirchler and Maciejovsky 2002).

Unlike most behavioral research on asset pricing, the research reported here uses experimental methods to examine the causes and nature of biases in pricing and probability judgment. We first develop a psychological model of judgment that encompasses both probability judgment and asset pricing as subjective measures of uncertainty and use this model to develop hypotheses about the pattern of biases expected to be found for these measures. We examine asset prices set in a simulated stock market setting where participants learn all cues and probability relationships through feedback and experience, and we evaluate how observed biases in probability judgment change when uncertainty is measured through the more familiar measure of pricing.

Our model is grounded in an influential conceptualization of judgment that distinguishes between two reasoning systems (Kahneman and Frederick 2002, Slovic 1996, Stanovich and West 2000; for a recent

Figure 1 Model of Case-Based Probability Judgment



review, see Evans 2008). A heuristic or intuitive system (System 1) provides a rapid, effortless assessment that is typically made with little conscious awareness. System 1 provides a heuristic assessment in the sense that it responds to a limited number of salient or easily evaluated attributes of the judgment object while neglecting other diagnostic information about the intended target. A deliberative system (System 2) provides a slower, effortful consideration of the fit between the judgment task and the output of System 1 and makes adjustments to that output as appropriate in arriving at a final judgment. In the absence of adequate correction by System 2, normatively important factors that affect outcome likelihood will be systematically neglected or underweighted because they are not captured in the initial, automatic evaluation of the heuristic attributes.

What normatively required considerations are likely to be neglected by the heuristic assessment process? The distinction between case-specific evidence and class-based factors—sometimes termed the “strength” versus “weight” of the evidence (Griffin and Tversky 1992)—helps to organize variables that have been found to be underused or neglected in studies of probability judgment (Brenner et al. 2005, Koehler et al. 2002, Tversky and Kahneman 1974). Case-specific evidence (e.g., a specific company’s price/earnings ratio) is what differentiates the case at hand, whose outcome is being assessed, from other cases drawn from the broader class or population. Class-based factors (e.g., overall economic conditions) are aggregate characteristics derived from the broader class or population. Results of many psychological studies indicate that intuitive judgments of probability tend to be primarily responsive to case-specific evidence (evidence strength) and largely neglect class-based factors (evidence weight) (see, e.g., Gilovich

and Griffin 2002, Griffin and Tversky 1992, Kahneman et al. 1982).

A simple schematic model of the case-based judgment process is presented in Figure 1. The top portion of the model represents the processing of individual case cues; the bottom portion of the model represents the processing of aggregate class cues—in particular, the overall base-rate (BR) frequency of a particular outcome and the diagnostic value of the cues in forecasting that outcome. When faced with the task of producing a probability judgment, System 1 generates an *impression* of the strength of the case at hand to produce a designated outcome (e.g., for a stock price to increase). This intuitive assessment is presumed to be based exclusively on case-specific evidence and, as indicated in Figure 1, can be measured through explicit instructions to rate the strength of evidence provided by the case-specific information. Many studies using support theory (e.g., Brenner et al. 2002, Fox 1999, Koehler 1996, Koehler et al. 1997, Tversky and Koehler 1994) have shown that rated strength or support for the specific case is highly predictive of reported probability, even when the weight of evidence is low. The purely case-based model illustrated in Figure 1 shows the limiting case where there is no System 2 adjustment of the impression generated by System 1; the case-based impression is simply transformed onto the probability scale for reporting. Thus, the class-based processing illustrated at the bottom of the figure has no impact on the subjective probability judgment for the designated case. This “neglect” model implies that System 2 fails to incorporate class-based factors that are relevant to the probability judgment task, even though those characteristics of the class environment may have been learned and can be reported in response to a query specifically addressing the class environment.

According to the purely case-based model, as class factors such as outcome base rate and cue diagnosticity change, we expect support ratings and probability judgments based on the same set of cues to remain essentially constant. Case-based judgment thus yields a specific pattern of shifting miscalibration across different combinations of base rate and cue diagnosticity precisely because the relation between case-based impression of support and judged probability remains invariant. This pattern of biased probability judgment has been observed in laboratory tasks of probability judgment (Griffin and Tversky 1992, Massey and Wu 2005), in real-world expert judgment tasks (Koehler et al. 2002), and in a stock market simulation where participants learned class conditions by experience and received monetary incentives based on the calibration of their probability judgments (Brenner et al. 2005). These patterns are derived and described in the context of the judgment model presented below.

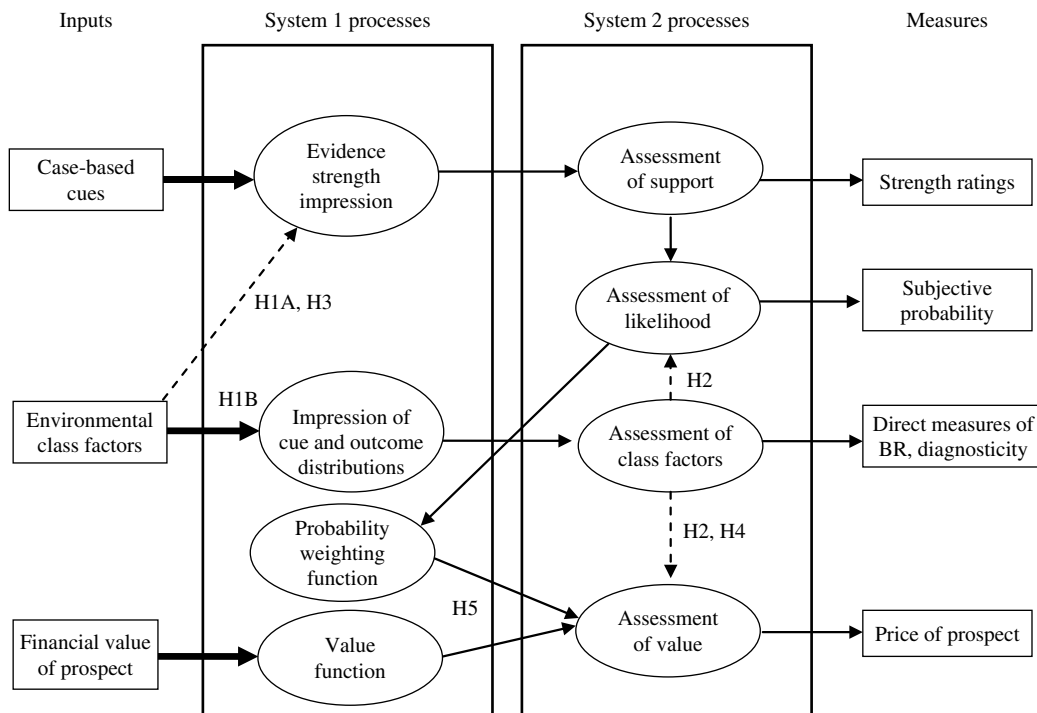
Subjective uncertainty can be measured by direct probability judgments or by the reservation price of a future asset, such as in prediction markets where an asset is worth some positive amount of money if the specified event occurs and nothing otherwise. Figure 2 presents the case-based pricing model, an expanded model of intuitive uncertainty assessment that incorporates the asset pricing task. This figure differs from Figure 1 in two ways: psychophysical functions transforming payoffs into subjective values and probabilities into decision weights are added to

the bottom of the figure, and additional paths are added to illustrate ways in which class-based factors may affect asset pricing.

Setting a price for an uncertain asset requires the integration of the subjective value of the payoff and the subjective likelihood of the event occurring. Following prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992), we incorporate two psychophysical functions into our framework: a value function that translates the objective payoff into subjective value and a probability weighting function that translates the likelihood assessment into a decision weight. We chose to characterize these functions as operations of System 1 in light of, among other things, experimental and neuropsychological evidence linking the degree of nonlinearity in these functions to reward encoding or affective evaluation processes (Hsee and Rottenstreich 2004, Hsu et al. 2009, Rottenstreich and Hsee 2001, Rick 2010), which are widely viewed as System 1 operations. The hypotheses we develop below, however, do not directly depend on this characterization.

In the case-based pricing model (Figure 2), the asset pricing task may differ from the probability judgment task because of the psychophysical transformations and because of changes in case-based processing. The System 1 operation of the prospect theory transformation functions may cause observed asset prices to deviate systematically from corresponding judgments of probability. As described more formally below, the typically shallow slope of the nonlinear probability

Figure 2 Model of Case-Based Probabilities and Prices (with Dashed Class Sensitivity Possibilities)



weighting function in the middle of the probability scale is predicted to blunt the effect of changes in evidence strength on asset prices relative to their impact on judged probabilities. The value function adds diminishing sensitivity in the evaluation of payoffs, which also alters the effect of probability changes on asset prices. Furthermore, loss aversion, the value function's steeper slope for losses than for gains, implies that measuring asset value through buying and selling prices will lead to divergent results. These effects are predicted to differentially affect asset pricing and probability judgment and should be largely independent of the effects of the class factors such as base rate and evidence diagnosticity.

The pricing task could also potentially lead to greater incorporation of class characteristics through either System 1 or System 2. The asset pricing task could trigger a System 1 heuristic evaluation that places greater weight on class-based factors. Because each judgment now carries direct consequences for the decision maker in the form of monetary gains or losses on each trial and hence greater hedonic impact, the case-based impression itself could come to incorporate better the overall base rate or diagnostic value. For example, a particularly low base rate of success and the associated failures to achieve a positive outcome could lower all case-based impressions. In effect, exposure to a "bear market" could begin to color the assessments of individual stocks. Under this "economic impact" hypothesis, class-based factors encountered during repeated asset pricing would directly influence the impression of evidence strength generated by the System 1 heuristic assessment process, and this effect would manifest itself in strength ratings as well as in asset prices.

The pricing task might also prompt greater System 2 adjustment for class-based factors than is found for judgments of probability. Even if System 1 generates an assessment that exclusively reflects the strength of case-specific evidence, System 2 could correct for or otherwise integrate considerations of class-based factors in setting a final price on the asset. There are several reasons to suppose that pricing tasks may lead to more System 2 intervention and correction for the class-based factors than is found for probability judgments. First, the greater hedonic impact of gains and losses on each trial may lead to a stronger memory trace for the base rate that may make it more accessible to System 2 use. Second, System 2 adjustment may be more natural in the pricing task as a result of the extensive experience people have in evaluating and setting prices. The general goal of valuation may prompt recognition of the impact of aggregate characteristics of the market in which the asset is traded, above and beyond any inherent value the asset itself might hold.

For example, home buyers may explicitly recognize that housing market conditions affect the price of a house above and beyond the features and condition of the house itself. Third, economic settings may encourage more explicit considerations of rationality constraints, whereby calculation of expected value overwhelms intuitive evaluations, and greater effort is expended than under direct probability assessment. Under all these possibilities, class-based factors could influence pricing judgments without necessarily influencing System 1 perceptions of evidence strength.

We summarize the various hypotheses suggested by the two-systems case-based framework in Table 1. Hypotheses 1A and 1B summarize the basic case-based assumptions supported by previous studies of probability judgment: System 1 automatically creates an index of the intuitive strength of evidence based on the predictive cues and monitors the outcome environment to assess the aggregate likelihood of the target outcomes. H1A states that ratings of evidence for a particular case are fully case based and insensitive to class considerations. H1B states that class information will be encoded and can be expressed in response to appropriate questions about the class aggregate properties. If both H1A and H1B hold, then class considerations are in fact detected and learned but may not be applied to judgments about a particular case. This is reasonable for judgments of evidence strength but leads to systematic miscalibration for judgments of probability (Griffin and Brenner 2004, Koehler et al. 2002). Previous research with a probability judgment task where relations between cues and outcomes in the environment are learned from experience found that outcome base rate was encoded accurately but the diagnosticity or weight of the cues was only weakly encoded in memory (Brenner et al. 2005).

Hypothesis 2 summarizes the possible qualitative levels of sensitivity to class considerations. Judgments may completely ignore class considerations (H2A); may be somewhat, but insufficiently, sensitive (H2B); or may be fully and appropriately sensitive (H2C). Brenner et al. (2005) found partial correction for base rate (supporting H2B) and no correction for diagnosticity (supporting H2A). In the presence of sensitivity to class factors, diagnosing whether correction is attributable to System 1 or 2 requires considering whether System 1 evaluations of the case-specific evidence strength show class sensitivity. In general, if H1A holds, and strength ratings show no class sensitivity, then any sensitivity to class manifested in H2B or H2C is attributable to System 2 correction. Any System 1 level correction would appear as class sensitivity in the strength ratings and would lead to rejection of H1A. In short, contributions of Systems 1 and 2 can be potentially teased apart by contrasting the class

Table 1 Hypotheses

System 1 sensitivity
H1A: Strength or support ratings are sensitive to the strength of evidential cues but unaffected by class-based factors (System 1 case sensitivity and specialization).
H1B: Direct ratings of base rate and cue diagnosticity are sensitive to the corresponding class factors (System 1 class sensitivity).
System 2 correction/class sensitivity in individual case judgments
H2A: Final judgments are unaffected by class-based factors (fully case-based judgment).
H2B: Final judgments show some (insufficient) sensitivity to class-based factors (partially case-based judgment).
H2C: Final judgments are fully corrected for class-based factors (fully Bayesian judgment).
Economic impact (asset pricing versus probability judgment)
H3: Strength ratings made in context of pricing judgments, in comparison to those in context of probability judgments, are more sensitive to class-based factors (enhanced System 1 sensitivity to class).
H4: Pricing judgments are more sensitive than are probability judgments to class-based factors (enhanced correction by System 2), beyond any enhanced sensitivity to class-based factors in strength ratings as captured by H3.
H5A: Selling price judgments are less sensitive to case-based cues than are probability judgments (due to System 1 psychophysics of valuation and uncertainty as captured by value and weighting functions of prospect theory).
H5B: Selling price judgments are less sensitive to case-based cues than are buying price judgments (due to System 1 psychophysics of valuation and uncertainty as captured by value and weighting functions of prospect theory).

sensitivity of the strength judgments and the probability/price judgments.

The new “economic impact” hypotheses tested in this paper are Hypotheses 3 through 5, which describe the three routes by which pricing judgments may differ from probability judgments. First, the pricing task may enhance System 1’s sensitivity to class-based factors in the process by which evidence strength is evaluated (H3). Second, there may be greater adjustment by System 2 for class-based factors when generating prices rather than probability judgments on the basis of perceived evidence strength (H4). Third, setting aside the question of sensitivity to class-based factors, prices are expected to be less sensitive than probability judgments are to variations in case-based evidence due to the nonlinear transformations used to translate objective value and probability into subjective value and decision weight (H5A and H5B).

There are, of course, additional testable predictions that follow from the two-systems case-based framework as depicted in Figure 2, particularly with regard to underlying cognitive processing. For instance, operations carried out by System 2 ought to be more susceptible to interference from working memory load and other resource-demanding tasks than should operations carried out by System 1, which are presumed to require few resources. In the present research, we chose to focus on the subset of predictions that directly concerned how pricing and probability judgment might differ in sensitivity to class-based factors.

The Asset Valuation Task and Case-Based Judgment Model

We test these hypotheses using a simulated stock market prediction task in which participants judge the likelihood of a company’s stock price either increasing or decreasing on the basis of several predictive

cues about the company. Four cues are presented in a bar graph for each company: domestic sales, domestic costs, foreign sales, and foreign costs. The sales cues are positively related to a subsequent stock price increase, and the cost cues are negatively related to a subsequent stock price increase. The diagnosticity or predictive value of the cues is varied across conditions, as is the overall base rate of price increases. The cues represent the strength of the case-based evidence, and our model of judgment is based on an overall aggregate measure of case-based evidence derived from the cues.

On each trial, experimental participants see the evidential cues for a given company and then learn the target outcome (stock price increase or decrease). After a series of learning trials in which the outcome base rate and the cue diagnosticity can be learned, participants encounter test trials in which they make judgments about whether the stock will increase or decrease. In some conditions, participants directly assess the probability of an increase. In other (pricing) conditions, they assign buying or selling prices to an asset that will be worth \$10 if the stock price increases and \$0 otherwise.

We model case-based judgment by regressing, across trials for each participant, the log-odds-transformed judgment (either probabilities, selling prices, or buying prices, depending on the task) on an aggregate measure of case-based evidence. The aggregate measure used is total sales (foreign and domestic) minus total costs. Other measures of case-based impression such as a weighted average of the cues or using multiple predictors, are possible. However, the notion of case-based judgment implies that the judge bases her judgment on a natural and immediate impression of case strength; therefore, a single summary predictor best captures the essence of case-based judgment. In terms of ability to predict the outcome,

little is lost by using the aggregate measure; it correlates very highly ($r = 0.98$) with the optimal weighting of the four cues.

We first consider a case-based judgment model for direct probability judgments and then expand to the case of buying and selling prices, incorporating value and probability weighting functions. Case-based judgment is instantiated by representing the log-odds (L_p) of the subjective probability that the stock price will increase as a linear function of the aggregate case-based evidence (C):

$$\ln\left(\frac{P}{1-P}\right) = L_p = \alpha + \beta C. \quad (1)$$

The dependent variable in our studies is the log-odds of the judgment, abbreviated L with a subscript tailored to the particular judgment for the task (probability, selling price, or buying price). Log-odds are used because the scale is unbounded and the ideal Bayesian likelihoods can be easily expressed in this form, allowing for crisp comparisons between the observed and optimal values. Sensitivity to class-based evidence manifests itself in the intercept (α) and slope (β) parameters in this model. For good calibration of judged likelihoods, changes in the base rate of the outcome should influence the intercept α , and changes in the diagnosticity of the cues should influence the slope β . Appendix A derives the optimal (Bayesian) intercept and slope, in terms of the cue distributions for increase and decrease trials:

$$\alpha^* = \left[\ln\left(\frac{P(\text{Inc})}{P(\text{Dec})}\right) - \frac{(\mu_1^2 - \mu_0^2)}{2\sigma^2} \right] = \ln\left(\frac{B}{1-B}\right) - \frac{d\mu}{\sigma},$$

$$\beta^* = \frac{\mu_1 - \mu_0}{\sigma^2} = \frac{d}{\sigma}.$$

The optimal intercept α^* depends on the log-odds-transformed base rate of increases B plus another term that adjusts for the slope if the cue distribution midpoint μ is nonzero. If the cue distributions are centered around $\mu = 0$, then the ideal intercept is simply the log-odds of success. The optimal slope depends on d , the diagnosticity of the cues, rescaled by the standard deviation σ of the conditional cue distributions.

The primary hypothesis of case-based judgment is that the observed parameters α and β will remain constant across different judgment environments despite changes in base rate B and diagnosticity d . A Bayesian judge needs to adjust α and β to match the optimal values α^* and β^* based on the class factors; a case-based judge maintains essentially the same judgment model parameters regardless of the class factors.

Figure 3 contrasts the predictions of purely case-based judgment and Bayesian judgment. The panels on the left depict the judgment model relating

subjective log-odds L_p to the case-based evidence C . The panels on the right depict the implied calibration curves given this judgment model and a particular environment defined by base rate B and evidence diagnosticity d .

The top panels show purely case-based judgment (H2A) in which the judgment model is constant—with unchanging α and β across different class (outcome base rate and evidence diagnosticity) conditions. This constant judgment model leads to calibration curves that vary systematically over class factors: overly high probability judgments (overprediction) when base rates are low and overly low probability judgments (underprediction) when base rates are high. Case-based judgment also implies overly extreme predictions resulting in a shallow slope (overextremity) when cue diagnosticity is low and insufficiently extreme predictions resulting in an overly steep slope (underextremity) when cue diagnosticity is high.

The bottom panels depict optimal, Bayesian judgment (H2C). Here, the judgment model varies with the class factors, according to the relationships shown above for α^* and β^* . The result is consistently good calibration across different levels of base rate and cue diagnosticity. A constant pattern of perfect calibration implies an adaptive and class-sensitive judgment model, whereas purely case-based judgment implies diverging patterns of miscalibration under different environmental conditions.

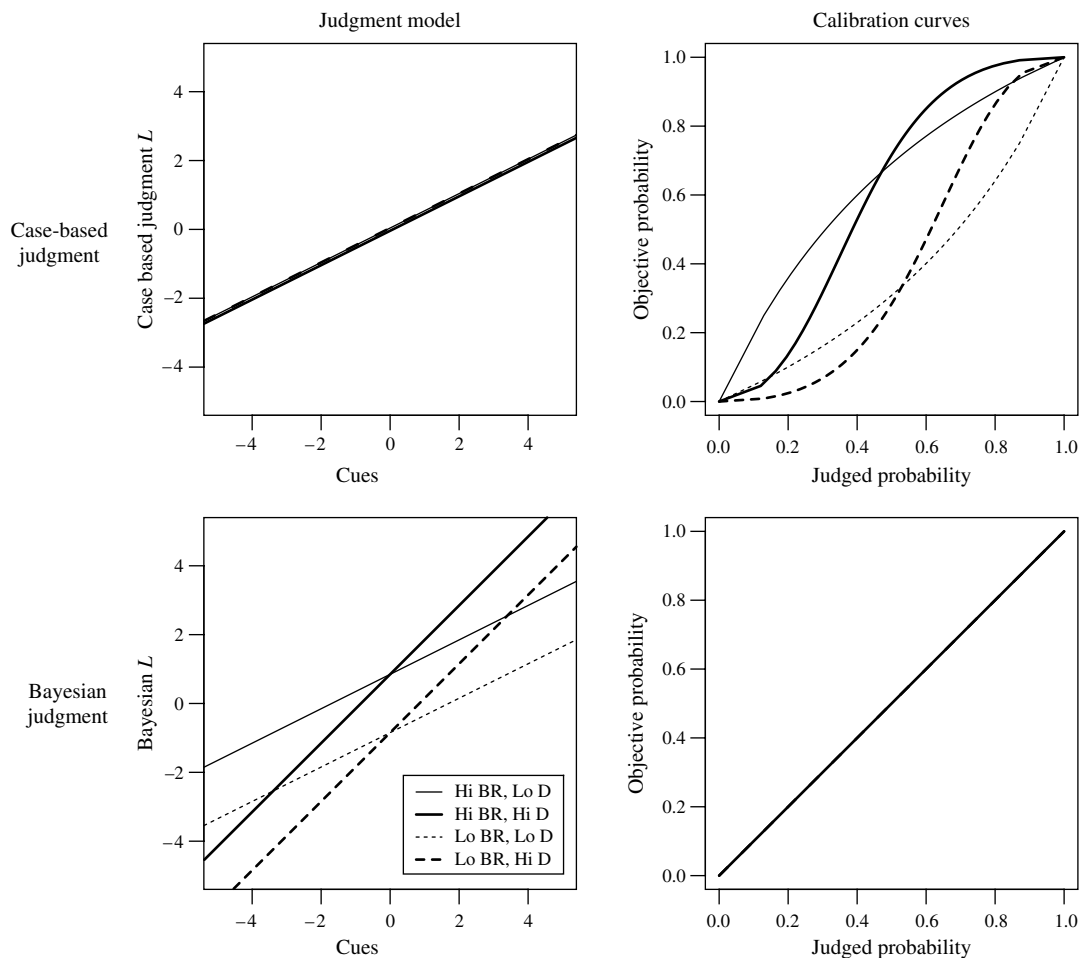
Setting Asset Prices

We extend our analysis to the task of determining the subjective value of an uncertain asset. Similar to the approach of Fox and Tversky (1998), we combine a model of subjective probability—the judgment model described above—with a model of decision making (prospect theory). Consider the task of setting a *selling price* S for an asset that pays \$1 for a success (stock price increase) and \$0 for a failure (stock price decrease). The selling price is determined so that the participant is indifferent between the subjective value of the sure amount S and the subjective value of the asset, which yields \$1 with subjective probability p and \$0 with subjective probability $1 - p$:

$$v(S) = w(p)v(1) + w(1-p)v(0).$$

Setting $v(0) = 0$ and $v(1) = 1$, we get $v(S) = w(p)$. We use the linear-in-log-odds judgment model (Equation (1)) to represent the subjective probability p . Now consider functional forms for the value function $v(\cdot)$ and probability weighting function $w(\cdot)$. We choose functions that are flexible enough to account for heterogeneity across judges but also allow clear insight into the qualitative differences between probability judgment and pricing.

Figure 3 Relation Between Judgment Models and Calibration Curves



Probability Weighting Function. We use the following two-parameter weighting function, discussed in some detail by Gonzalez and Wu (1999):

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma}.$$

The δ parameter allows for elevation changes and the γ parameter allows for slope/sensitivity changes. The typical empirical result, illustrating relative insensitivity to changes in moderate probabilities, is $\gamma < 1$, which represents the traditional inverse-S-shaped weighting function with sharp sensitivity near the extremes (0 and 1) and relative insensitivity elsewhere. Conveniently, this function is a linear function of the log-odds of subjective probability: $L_w = \ln \delta + \gamma L_p$.

Value Function. For the value function, we use a one-parameter function that is also linear in log-odds:

$$v(x) = \frac{\theta x}{\theta x + (1-x)}.$$

This function is sufficient for modeling risk aversion ($\theta > 1$), risk neutrality ($\theta = 1$), and risk seeking

($\theta < 1$) within the relevant range of outcomes between 0 and 1. The linear in log-odds form for the value function fits well with the specifications of both the probability judgment model (case-based judgment, linear in log-odds) and the probability weighting function (log-odds of the weights $w(\cdot)$ are linear with the log-odds of p). This function is also approximated quite well by the commonly used power function.

Given the relationship between selling prices and subjective probability $v(S) = w(p)$, we apply a logit transformation, which yields simple linear expressions for each side of the equation:

$$\ln(\theta) + L_S = \ln \delta + \gamma L_p.$$

Inserting the model of probability judgment $L_p = \alpha + \beta C$ yields a final expression for the log-odds of the selling prices:

$$L_S = (\ln \delta - \ln \theta + \alpha \gamma) + (\beta \gamma) C.$$

In this framework, log-odds-transformed selling prices will be a linear function of the aggregate cues (C), with the slope for probability judgments β

multiplied by the weighting function parameter γ . Because the γ parameter is typically less than 1 (for inverse-S-shaped curvature of the weighting function), we expect that the relationship between selling prices and case-specific evidence will be dampened relative to the relationship between probability judgments and case-specific evidence (H5A).

Note also that the intercept for selling prices ($\ln \delta - \ln \theta + \alpha \gamma$) is different from the intercept for probability judgments (α). In terms of predictable qualitative changes, we expect risk aversion to depress judgments based on the $-\ln \theta$ term. The flatness of the weighting function suggests a dampening of the intercept from α to $\gamma \alpha$, but this could yield either an overall increase or a decrease depending on the sign of α . Also the $\ln \delta$ term could either increase or decrease the intercept depending on whether $\delta > 1$ or $\delta < 1$. In contrast to the slopes, there is not a clear directional prediction to make regarding the comparison of the intercepts in pricing and probability.

Overview of Study Procedure

In a computer simulation of a stock market environment, participants assigned prices to assets whose value depended on directional changes in a company's stock price. Specifically, participants were told that they held (or could buy, in Study 3) a certificate that would be worth \$10 if the associated company's stock price increased the next quarter but would be worth \$0 if the stock price decreased. The value of the asset was therefore uncertain when the price was set and implicitly depended on the probability of a stock price increase.

For each asset, participants received case-specific sales and cost information (in the form of a bar graph) about the company in question. There were four cues: domestic sales, domestic costs, foreign sales, and foreign costs. Higher sales and lower costs were associated with higher probabilities of stock price increases. The magnitude of this association (i.e., the diagnostic value of the cues) could be learned from experience, as could the overall prevalence (i.e., base rate) of stock price increases in the market.

Participants were presented with a series of companies and for each were asked to assign a price to an asset that paid \$10 if the company's stock price increased the following quarter and \$0 otherwise. Responses were made on a \$0 to \$10 scale, separated into 25-cent intervals. Participants were paid based on the accuracy of their pricing judgments using the Becker–DeGroot–Marschak incentive-compatible payoff scheme (Becker et al. 1964). The specific implementation of the payoff scheme varied from study to study, as described below.

The diagnostic value of the cues and the base rate of increasing stock prices was manipulated between subjects. Cue diagnosticity was varied as follows. Each

cue value (domestic sales and costs, foreign sales and costs), conditioned on whether it was associated with a stock price increase or decrease, was represented as a normally distributed variable with unit variance. The sales cue distributions for companies with increasing stock prices had a higher mean than those for companies with decreasing stock prices, and vice versa for the costs cue distributions. The degree of separation between conditional cue distributions for companies with increasing and decreasing stock prices determines the diagnostic value of each cue. For both sales and costs, the diagnostic value of domestic indicator cues was set to be higher than that of foreign indicator cues. In the low diagnosticity (low D) condition, the separation between increasing and decreasing cue distributions was 0.8 standard deviations (SDs) for domestic indicator cues and 0.4 SDs for foreign indicator cues. In the high diagnosticity (high D) condition, the separation was larger: 1.2 SDs for domestic indicator cues and 0.8 SDs for foreign indicator cues. Based on these generating distributions, if a simple aggregate “company performance” measure is generated from the four cues by adding total sales and subtracting total costs, the resulting value correlates 0.50 with the dichotomous outcome variable in the low D conditions and 0.68 in the high D conditions.

The base rate of stock price increases (i.e., overall “bullishness” or “bearishness” of the market) was also varied between subjects. Across different studies, in the low base-rate (low BR) condition, 30% or 40% of companies (depending on study) had stock price increases, whereas in the high base-rate (high BR) condition, 70% of companies had stock price increases.

The set of companies associated with a particular condition was constructed by first setting the proportion of companies with increasing stock prices according to the desired base rate and then sampling cue values for each company from the appropriate conditional distribution. Across different studies, there are small differences in the class characteristics due to sampling variability.

Following the pricing task, participants rated the strength of evidence provided for 10 randomly chosen companies (allowing tests of H1A). Participants also rated the perceived outcome base rate and evidence diagnosticity across all trials (allowing tests of H1B).

Study 1: Pricing vs. Probability Judgments

The first study contrasts performance in the experimental asset pricing task with performance in a probability judgment task. One group of participants assigned a price (on a \$0 to \$10 scale) to an uncertain asset paying \$10 if the company's stock price

increased (and \$0 if it decreased). Another group of participants directly assessed the probability that the stock price would increase, using a 0% to 100% scale.

This study allows tests of many of the hypotheses in Table 1, most notably H4. According to the economic impact hypothesis, sensitivity to class factors will be enhanced in the asset pricing task, relative to that in the probability judgment task, because of the greater familiarity and hedonic impact of the response scale.

Method

Participants. Participants were 136 business students at the University of British Columbia and the University of Florida. Data from five participants were dropped from the analysis because they did not use the cues appropriately, as evidenced by outlying negative correlations between judged probability or price and the cues. There were 74 remaining participants in the pricing task and 57 in the probability judgment task.

Design. The study independently varied the base rate of stock price increase (30% versus 70%), the level of cue diagnosticity (aggregate cue-outcome correlations of 0.49 versus 0.65), and the evaluation task (pricing versus probability judgment). Participants were randomly assigned to one of the eight possible combinations of base rate, cue diagnosticity, and task.

Participants made a pricing or probability judgment for 100 trials (companies), each of which was followed by immediate outcome feedback indicating whether the company's stock price had increased or decreased. The first 20 trials were treated as practice in a separate block labeled as a "training session," for which there were no real payoffs. The remaining 80 trials constituted the second, "test" block, for which real monetary payoffs were offered. One trial from this block was chosen at random at the end of the experiment and was played for real money using the Becker-DeGroot-Marschak procedure described below.

Procedure. Participants completed the task on a computer using a standard web browser. They completed the task independently on individual computers during group sessions consisting of 15 to 20 participants.

Pricing Judgments. Participants were asked play the role of a securities analyst who has just begun to follow the Stockholm Stock Exchange. They were told that their first task upon joining the short-term certificates pricing desk was to complete a training session in which they would learn to predict which companies' stock prices will go up between one quarter and the next. They were told that they would see a series of 20 companies whose performance would be displayed in graphs of their domestic and foreign

costs and sales during that quarter, and would later discover whether the company's stock price increased or decreased the following quarter.

Participants were told that their brokerage held short-term certificates tied to each company's stock price change. The certificate would pay \$10 if the stock price increased but would be worthless if the stock price decreased. Participants were told that their task was to set the lowest price at which they would be willing to sell each certificate. It was pointed out that a higher price indicated a greater perceived chance that the certificate will pay off.

Following the 20-trial training session, participants were told that they would now be setting prices with real monetary outcomes. They were told that they would see 80 more companies drawn from the same market as the 20 encountered in the training session. Again, their task would be to set a selling price on a certificate that paid \$10 if the company's stock price increased. They were told that one of the 80 trials would be selected at random and the outcome played for real money.

The Becker-DeGroot-Marschak procedure for determining the payoff from the randomly selected trial was then described as follows. The computer would draw a random market clearing price between \$0 and \$10. If the selling price set by the participant was less than the market price, she would sell the certificate and receive the market price. If the selling price was greater than the market price, she would hold the certificate and be paid either \$10 or \$0 depending on whether the stock price increased or decreased. Participants were told that this procedure made it in their best interest to honestly report the minimum price at which they would be willing to sell each certificate and that there was no benefit in stating anything other than their true selling price. An example was provided in which setting a price higher than the individual was actually willing to accept could lead to an undesirable outcome with a missed opportunity to sell the certificate at an acceptable clearing price. In summary, participants were told, "If the price you state is too high or too low, then you are passing up opportunities that you would prefer."

On each trial, participants were presented with a company, depicted by bar graphs representing foreign and domestic sales and costs. They were told, "You hold a short-term certificate in Company [number]. If the stock price *increases*, the certificate is worth \$10; but if the stock price *decreases*, it will pay \$0. *What is the minimum price at which you would be willing to sell this certificate?*" Participants indicated their minimum selling price by clicking on a \$0 to \$10 scale with labeled whole-dollar increments, within each of

which there were three sequential plus signs representing 25-cent sub-increments. Participants then saw an outcome feedback screen indicating their selling price, whether the company's stock price actually increased or decreased, and consequently whether the certificate paid \$10 or \$0.

Probability Judgments. Participants assigned to the probability judgment task evaluated the same set of 100 companies as did those assigned to the pricing task. No mention was made of short-term certificates, however, and participants were asked only to judge the probability that the target company's stock price would increase the following quarter. Participants made their judgments on a scale that was structurally identical to that used in the pricing task, ranging from 0% to 100% in labeled 10% increments, each of which was separated by three further, unlabeled sub-increments.

As with the pricing task, the first 20 trials were said to constitute a training session, while the last 80 trials involved real monetary outcomes. Specifically, participants were told that a measure of the accuracy of their judgments would be calculated for these 80 trials and they would receive a payment based on their accuracy. As in the pricing task, the payoff scheme was said to be such that it was in participants' best interest to report their true judgment, in this case, of the probability of a stock price increase.

Strength Ratings. Following the main pricing or probability judgment task, participants were presented with 10 company profiles, drawn from the set of profiles seen earlier. Participants were asked to rate the "strength" of each company's financial performance based on the sales and costs cues. Ratings were made on a 0 ("very weak") to 10 ("very strong") scale. We take these ratings as a measure of the strength of the impression conveyed by the case-based information characterizing each company. These ratings can be used to evaluate H1A as well as determine whether any improvement in class sensitivity is attributable to System 1 or System 2 processes.

Base Rate and Diagnosticity Estimates. Finally, to test whether participants were sensitive to features of the judgment environment when directly asked (H1B), participants were asked to provide two estimates based on their experience with the 100 companies they evaluated during the pricing or probability judgment task. First, they were asked to estimate how many of the 100 companies had experienced a stock-price increase. We take this as a memory-based estimate of the base rate of stock price increases. Second, they were asked to estimate for how many of the 100 companies they had correctly anticipated whether its stock price would increase or decrease. We use this as a measure of the perceived diagnosticity of the

cues, that is, how well they could be used to predict the change in a company's stock price.

Results

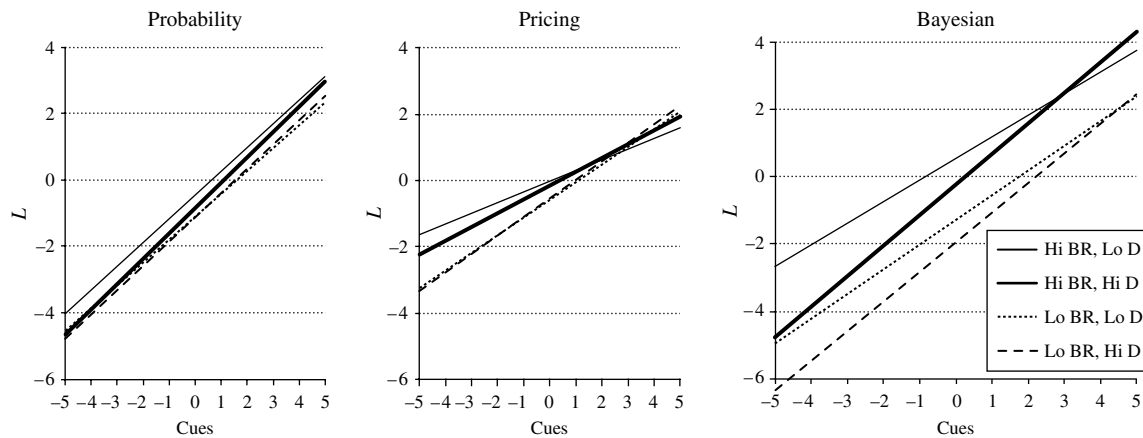
Direct Judgments of Aggregate Class Characteristics. First we consider the judgments of aggregate class characteristics, which provide tests of H1B. Overall judgments of stock price increases were extremely accurate and were higher in high base-rate conditions ($M = 72.9$) than in low base-rate conditions ($M = 29.6$, $F(1, 127) = 303$, $p < 0.0001$). Base-rate estimates were somewhat higher in the probability group ($M = 76.3$ and $M = 34.0$ in the high and low BR conditions, respectively) than in the pricing group ($M = 69.2$ and $M = 27.4$, $F(1, 127) = 8.0$, $p < 0.01$), but there was no interaction between type of judgment and base rate. In these direct judgments of aggregate class characteristics, participants were appropriately sensitive to different base rates.

Judgments of diagnosticity showed small differences in the appropriate direction (subjective accuracy of $M = 72.5$ and $M = 60.9$ for high diagnosticity in the probability and pricing groups, respectively, versus $M = 67.6$ and $M = 55.3$ for the low diagnosticity conditions), but the difference was not statistically significant ($F(1, 127) = 1.9$, $p = 0.18$). There was a main effect of judgment type ($F(1, 127) = 9.4$, $p < 0.01$), however, with the perceived accuracy judgments being systematically higher for the probability group.

Sensitivity of Strength Ratings to Case and Class Factors. Log-odds-transformed strength ratings were regressed on aggregate case evidence (C) for the 10 companies evaluated at the end of the study. The resulting intercepts and slopes give an indication of the extent to which class factors have been incorporated into the System 1 impression of a company's strength. Consistent with the first part of H1A, strength ratings were indeed strongly related to the case-based cues (average slope = 0.50, ranging from 0.47 to 0.52 across different base-rate and diagnosticity conditions). Consistent with the second part of H1A, there were no significant differences in either intercept or slope across class characteristics (intercept $F(3, 123) = 0.52$, slope $F(3, 123) = 0.29$). The strength ratings can be characterized as fully case based. This held equally for participants who had completed the pricing task and for those who had completed the probability judgment task (intercept $F(3, 123) = 0.34$, slope $F(3, 123) = 0.27$), indicating no enhanced System 1 sensitivity to class-based factors in the context of pricing (contrary to H3).

One unexpected result is that the average slope is marginally higher in the probability task ($M = 0.55$) than the pricing task ($M = 0.45$, $F(1, 123) = 3.6$, $p = 0.06$); the company strength ratings are somewhat more closely linked to the cues after making probability judgments than after pricing uncertain assets.

Figure 4 Average Judgment Models and Bayesian Standard for Probability and Pricing Task, Study 1



Sensitivity of Probability and Price Judgments to Class Factors. We turn now to the price and probability judgments. Over the 80 test trials, we estimated within-subject regressions predicting the log-odds of the target judgment (probability or price) from the aggregate cue summary (C). Figure 4 displays the resulting average judgment model for the probability and pricing tasks and also displays the ideal Bayesian judgment model for comparison. The primary qualitative comparison to note is the much greater change in the Bayesian models across conditions compared with the observed judgment models. These comparisons are elaborated below.

Sensitivity to Base-Rate Changes. First, consider the intercepts from these regressions. There was some sensitivity to base-rate changes, with higher intercepts for the high BR conditions ($Adj. M = -0.35$) than the low BR conditions ($Adj. M = -0.81$, $F(1, 123) = 6.6$, $p = 0.011$). There was also a main effect of the judgment type, with higher average intercepts for the pricing task ($Adj. M = -0.32$) than the probability judgment task ($Adj. M = -0.84$, $F(1, 123) = 8.8$, $p = 0.0036$). However, both probability and price judgments were equally sensitive to the changes in base rate; there were no interactions between the class factors and the task type. Crucially, there is no evidence of greater sensitivity for the pricing judgments, contrary to H4.

The observed sensitivity to base rate was markedly less than would be expected under Bayesian judgment, where the optimal intercepts are -1.61 and 0.16 for the low and high BR conditions, respectively. The observed sensitivity—a shift of 0.46 log-odds units—is a bit more than a quarter of the sensitivity (1.77) required for Bayesian judgment. The base-rate results are best characterized by H2B—minimal but insufficient sensitivity to base rate.

The insufficient sensitivity to base rate can be simply illustrated by comparing the average judgment

across base-rate conditions. In the probability judgment task, the average judged probabilities were 37.5% and 56.5% in the low (30%) and high (70%) base-rate conditions, respectively. In the pricing condition, the average prices were $\$4.25$ and $\$5.71$.

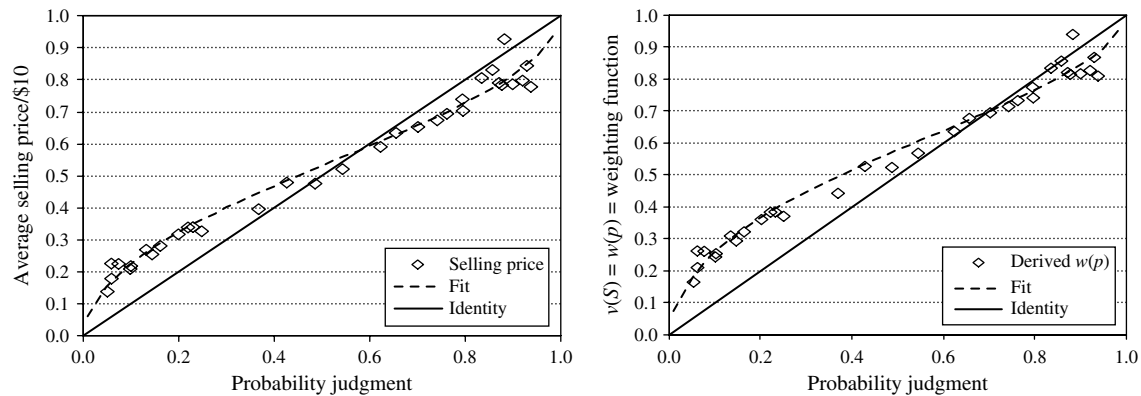
The difference between judged probability/price and the Bayesian ideal (“base-rate neglect”) is significant across base-rate condition ($F(1, 127) = 127$, $p < 0.001$), with no interaction between base-rate condition and task ($F(1, 127) = 1.18$, $p = 0.28$). For both tasks, judgments were too high when base rates were low and too low when base rates were high.

Note that the insensitivity to base rate appears in the trial by trial pricing and probability judgments despite the very accurate aggregate judgments of outcome base rate made at the end of the task. Participants can recognize the different base rates in their aggregate assessments, but individual judgments are largely case based and show only minimal sensitivity to base rate.

Sensitivity to Diagnosticity of Evidence. The estimated slopes from the within-subject cue regressions measure sensitivity to the case-based evidence. Appropriate sensitivity to cue diagnosticity would be manifested by steeper slopes in the high diagnosticity than in the low diagnosticity conditions. However, the average slope was virtually identical under high diagnosticity ($Adj. M = 0.64$) and under low diagnosticity ($Adj. M = 0.60$, $F(1, 123) = 0.52$, $p = 0.47$) and did not interact with task type ($F = 0.39$, $p = 0.53$). For comparison, Bayesian judgment entails an optimal slope of 0.69 for low and 0.89 for high diagnosticity. Overall, the slope results are consistent with H2A (no sensitivity to diagnosticity changes) and inconsistent with H4 (no improvement in pricing).

Differences Between Probability and Pricing. The dampening effect of the weighting function implies decreased sensitivity to cues for pricing relative to that for probability (H5A). The results support this

Figure 5 Relationship Between Selling Prices and Judged Probabilities



Note. Average selling prices versus average judged probability (left panel) and estimated probability weighting function given mild risk aversion (right panel).

hypothesis, with systematically shallower slopes for pricing ($Adj. M = 0.49$) than probability ($Adj. M = 0.74$, $F(1, 123) = 18.3$, $p < 0.001$).

Recall that H5A is based on the prospect theory probability weighting function blunting the sensitivity of selling price to changes in subjective probability. Given that we have data on both judged probability and selling prices in this experiment, we can attempt to reconstruct the probability weighting function. First, we divide the possible cue values into bins (e.g., -0.2 to 0.2 , 0.2 to 0.6 , etc.) For each bin of cue values, we calculate the average probability judgment (from the probability condition) and the average selling price (from the pricing condition). The average selling price is plotted against the average probability for each bin of cue values in the left panel of Figure 5. This figure shows an inverse S-shape similar to the traditional probability weighting function, most notably with relative insensitivity to changes in the middle of the probability scale. The graph also depicts risk seeking for low probabilities (i.e., selling prices are somewhat higher than expected value for low probabilities) and risk aversion for higher probabilities.

This graph does not fully isolate the weighting function, in that the assigned selling prices represent a combination of the weighting function and a subjective value function. The left panel of Figure 5 depicts the weighting function assuming a linear (risk neutral) value function. Incorporating mild risk aversion in the value function ($\theta = 1.2$) yields the weighting function depicted in the right panel of Figure 5. Note that in both panels, the function is approximated quite well by the fitted Gonzalez and Wu (1999) linear in log-odds decision weight function.

Summary. The results of Experiment 1 show no trial by trial sensitivity to evidence diagnosticity (H2A) and inadequate sensitivity to outcome base rate of success (H2B) in both the probability and pricing tasks. Comparing probability and pricing does not

support H4, showing no improvement in sensitivity to either base rate or diagnosticity in the pricing task relative to that in the probability judgment task, but supports H5A, with generally lower sensitivity to case evidence (shallower slopes) for pricing than probability, consistent with a flat-in-the-middle probability weighting function.

Replication. Although the Becker–DeGroot–Marschak payoff scheme was described in detail to participants prior to their completing the pricing task in Study 1, it was only implemented once (for a randomly selected trial) at the end of the experiment. Given this procedure, it is possible that the economic impact of the feedback was slight. In addition, participants might have thought, incorrectly and contrary to instructions, that there was an advantage in setting prices in a strategic manner that did not necessarily reflect their true valuation of each asset. Furthermore, having the payment contingent on a single trial may have altered strategies, perhaps by encouraging attention to certain trials that subjects may have thought to be the key payoff trial, or might have more generally affected attitudes toward risk. To address these concerns, we conducted a replication study in which the BDM payoff scheme was implemented on every trial. After participants ($N = 95$) set their price for a given certificate, a random market clearing price was determined and the appropriate transaction carried out and reported to participants. Participants were paid on the basis of their total earnings across the test trials. Despite these changes, the results were essentially identical to those reported for Study 1.

Study 2: Extended Experience

One might ask whether the pricing biases observed in the previous study, which we attribute to a focus on case-specific information and neglect of class-based considerations, would be reduced if participants were

given more extensive experience setting prices. The supplementary ratings collected in the previous study indicate that participants were given sufficient experience with which to accurately evaluate the relevant outcome base rate of success when explicitly asked to do so. Recognition of the relevance of this information to the pricing task, however, may require more experience performing the task in a particular market environment than was provided in the previous study. Further, given the relative insensitivity of the direct diagnosticity judgments to the diagnosticity manipulation, it is possible that extended experience is required before participants are able to reliably detect the difference in cue diagnosticity across conditions. To test these possibilities, participants in Study 2 completed an extended version of the pricing task with three times as many test trials.

Method

Participants. Participants were 112 undergraduates at the University of Waterloo. Data from four participants were dropped from the analysis because they did not use the cues appropriately, as evidenced by outlying negative correlations between judged price and outcome in the test trials.

Design and Procedure. The design and procedure were similar to that of the pricing condition of Study 1, except that the total number of trials was increased to 260, the first 20 of which were again treated as practice. The remaining 240 trials were divided into three blocks of 80 trials each. Participants were invited to take a short break between blocks if they wished.

The cue distributions were also somewhat different. The low and high base rates were 40% and 70%, respectively. Cue diagnosticity ranged from $r = 0.50$ to $r = 0.80$.

Another procedural change involved the response scale on which the pricing judgments were made. To make even clearer the implications of setting a particular minimum selling price, when the participant clicked on a value it divided the scale into two regions. The region with prices below the selected selling price was labeled the “hold zone,” and participants were instructed that if the randomly selected clearing price fell within that region (i.e., below their selling price), they would hold the certificate and receive either \$10 or \$0 depending on whether the company’s stock price increased or decreased. The region above the selling price was labeled the “sell zone,” and participants were instructed that if the randomly selected clearing price fell within that region (i.e., above their selling price), they would sell the certificate at the clearing price.

After participants clicked on a value and saw the scale divided into the hold and sell zones, they could

either change or confirm their selected price. Once a price was confirmed, they were informed of the clearing price and received feedback regarding whether the stock price had increased or decreased. For example, a participant who set a selling price that turned out to be lower than the clearing price might be told, in addition to whether the company’s stock price increased or decreased, “You offered to sell your Company #25 certificate for \$6.50. The randomly drawn clearing price was \$8.49. Therefore, you *sold* your certificate for \$8.49.” Alternatively, if the selling price was higher than the clearing price, they were informed of the clearing price, told that as a result they retained the certificate, and learned whether they received \$10 or \$0 as a result of holding the certificate. Clearing price information (and the associated hold zone and sell zone markers) were introduced only after the initial training trials in which participants set a price on each trial but received only outcome feedback regarding whether the company’s stock price increased or decreased. Participants received payment at the end of the experiment equal to their average proceeds across the full set of test trials.

Results

Differences in both base rate and diagnosticity were detected in the questions about class factors at the end of the study. Overall judgments of base rate were again remarkably accurate: higher in high base-rate conditions ($M = 68.6$) than in low base-rate conditions ($M = 41.1$, $F(1, 104) = 122$, $p < 0.0001$). Overall judgments of diagnosticity were higher in the high diagnosticity conditions ($M = 78.0$) than in the low diagnosticity conditions ($M = 66.8$, $F(1, 104) = 9.6$, $p < 0.01$). Thus, with extended learning trials, H1B is supported for both base-rate and diagnosticity judgments; participants were indeed sensitive to the class factors in their aggregate judgments.

Did the awareness of these class factors affect the trial by trial pricing judgments for the individual companies? Across all test trials, we fit judgment models predicting log-odds transformed prices (L) from the cue summary index (C) for each participant and display the average estimated parameters in Figure 7. Judgment models showed some sensitivity to base rate; the average intercept is higher for the high BR than for the low BR conditions (0.16 versus -0.33 , $F(1, 104) = 9.3$, $p < 0.01$). This sensitivity, however, was less than half of what would be required for Bayesian judgment (0.5 units observed change versus 1.2 required change). As in the previous study, sensitivity to base rate is best characterized by H2B. There was no significant effect of diagnosticity on the estimated slope of the judgment models, $F < 1$, consistent with H2A.

Figure 6 Average Judgment Models and Bayesian Standard for Pricing Task, Study 2

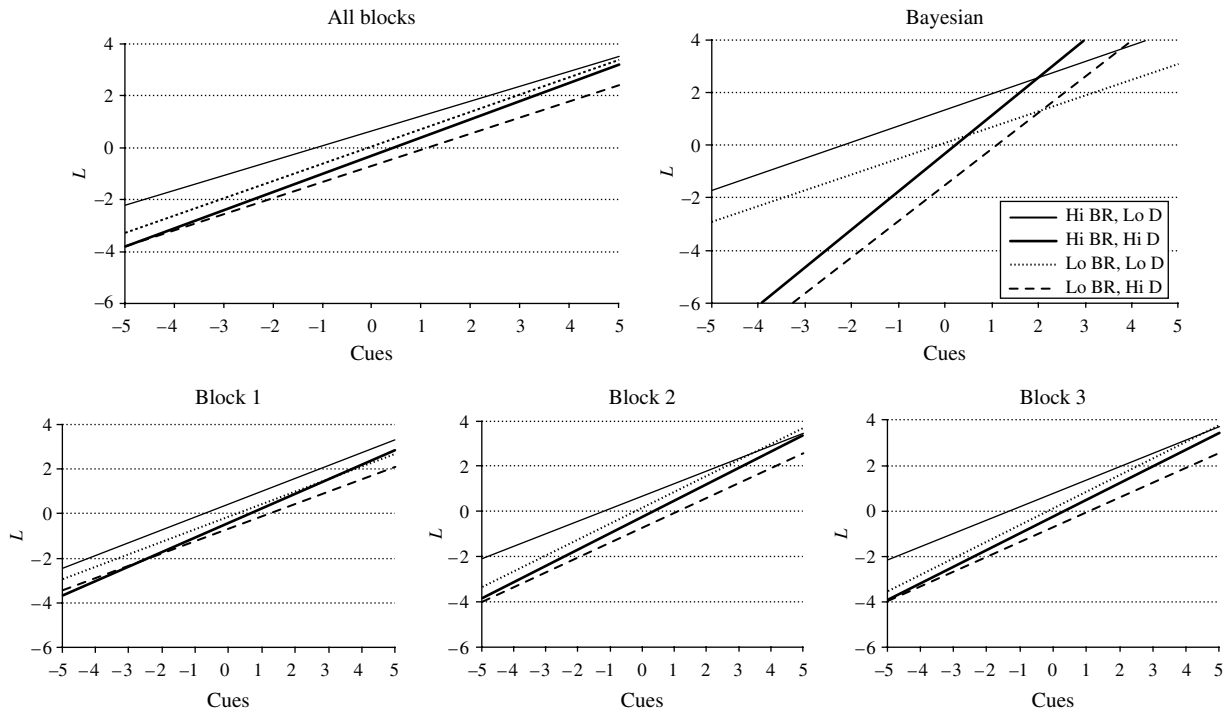


Figure 6 displays the average parameter values for three blocks of 80 trials each. There are no statistically significant effects of block on either the intercept or slope of the judgment model. In each block there is noticeable, but insufficient, sensitivity to base rate, which remains roughly constant across blocks. There is also no appreciable sensitivity of the slope to diagnosticity condition, and this is also stable across blocks. The results are well characterized by H2B for base rate and H2A for diagnosticity, with no indication that sensitivity to class factors increases with extended experience in the pricing task environment.

Study 3: Buying vs. Selling Prices

Participants in the previous studies set the minimum price at which they would be willing to sell each asset. Because they set selling prices, participants could not lose money in these studies—they either broke even or gained money on each trial. Study 3 introduces a condition in which participants set buying rather than selling prices. On each trial, buyers indicated the maximum price they would be willing to pay to obtain the asset. In contrast to the sellers, buyers face the possibility of losing money, which would happen if they bought an asset that did not pay off. There is substantial evidence that losses exert a larger impact on judgments and decisions than do comparable gains, a principle known as loss aversion (Kahneman and Tversky 1979, Tversky and Kahneman 1991). It is possible that the greater impact of potential losses faced

by buyers, relative to comparable gains faced by sellers, could enhance learning from experience, leading to greater sensitivity to class-based considerations such as outcome base rate and evidence diagnosticity. Put differently, it is possible that the sellers in the previous studies were complacent because of the consistent receipt of gains on each trial and as such were not strongly motivated to find ways to increase those gains. The threat of losses faced by buyers might provide stronger motivation to set unbiased prices.

In Appendix B, we describe a model for setting buying prices. This model yields two qualitative predictions contrasting buying and selling prices. First is the familiar prediction based on loss aversion and found in studies of the endowment effect (Kahneman et al. 1991): buying prices are expected to be systematically lower than selling prices, despite the use of an incentive-compatible elicitation procedure.

A more novel prediction is described in H5B—greater sensitivity to case-based evidence in buying prices than in selling prices. The mathematical details are presented in the appendix, but broadly speaking the source of the greater sensitivity for buying prices is the value function's greater steepness near the reference point separating the domains of gains and losses. Because the task of setting buying prices entails considering a range of possible outcomes straddling the reference point (e.g., a certificate bought for \$3 could result in an overall \$3 loss or a \$7 gain), that range of outcomes yields a wider spread of subjective values

given the greater sensitivity of the value function near the reference point.

Method

Participants. Participants were 294 business students at the University of British Columbia and the University of Florida. Data from 22 participants were dropped from the analysis because they did not use the cues appropriately, as evidenced by outlying negative correlations between judged probability and outcome in the judgment trials.

Design. The study had a two (base rate; low 40% versus high 70%) by two (diagnosticity; low $r = 0.5$ versus high $r = 0.8$) by two (role; buyer versus seller) full-factorial design. There were 129 participants in the buying condition and 143 in the selling condition.

Procedure. As in the previous studies, participants completed an initial 20 practice trials in which they set prices and received outcome feedback but did not see a randomly determined clearing price or the resulting transaction. This was followed by 80 test trials in which the clearing-price mechanism and payoff scheme were implemented. Procedure for the sellers was nearly identical to that of Study 2, in which the pricing response scale was divided into a hold zone and a sell zone. The only change was that an additional table on the outcome screen provided a summary of the financial outcome on the current trial along with a cumulative balance across all the test trials up to that point.

Buyers received instructions that were nearly identical to those of the sellers, except they were told that their brokerage could bid on a certificate (that would pay either \$0 or \$10) tied to each company's stock price. They were instructed to indicate the highest price that they would be willing to pay to buy each certificate. After the initial 20 practice trials setting buying prices, the clearing-price mechanism and payoff scheme were introduced. After setting their buying price, a clearing price was randomly determined. If the clearing price was less than the buying price set by the participant, the certificate would be purchased at the clearing price and the participant would gain either \$10 or \$0 (less the price at which the certificate was purchased) depending on whether the company's stock price increased or decreased. If the clearing price was greater than the buying price set by the participant, no transaction took place, but the participant was still informed whether the company's stock price increased or decreased and the resulting payoff value of the certificate. When the participant clicked on a buying price, the response scale was divided into a "buy zone" at or below the selected price and a "pass zone" above the selected price, indicating the consequences if the randomly selected clearing price fell in

either zone. As in the selling condition, buyers could either confirm their selected price or change it. It was pointed out, and illustrated with an example, that the payoff scheme made it in the participants' best interest to give their true buying price for each certificate.

Sellers began with a balance of \$0 in their brokerage account, to which they could only add money as a result of their transactions. Buyers began with a balance of \$100, to which money could be added or subtracted as a result of their transactions. Buyers were given the positive starting balance to avoid any changes in pricing strategy that might result from being in a deficit position and so that their final balance would be roughly equivalent to that of the sellers. Both groups were provided with a table showing their current balance after each pricing trial and were paid in proportion to their final balance. Average payment was \$2.48 for buyers and \$5.29 for sellers.

Results

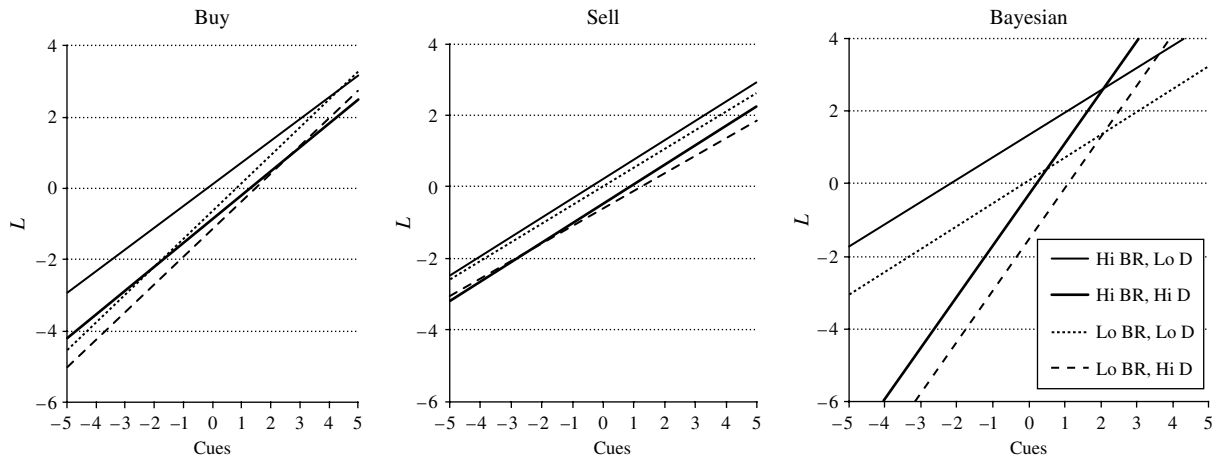
Direct Judgments of Aggregate Class Characteristics. In support of H2B, overall judgments of base rate were again quite accurate: higher in high base rate conditions ($M = 68.8$) than in low base rate conditions ($M = 44.0$, $F(1, 268) = 207$, $p < 0.0001$). Sensitivity to base rate was somewhat stronger for those setting buying prices (42.5 versus 71.1) than for those setting selling prices (45.4 versus 66.5, $F(1, 268) = 4.7$, $p < 0.05$).

We also find differences in aggregate judgments of the cue diagnosticity, especially for buying prices. Overall, participants judged they had accurately anticipated the outcome of 60.4% of the trials when diagnosticity was low versus 73.3% when diagnosticity was high ($F(1, 268) = 29$, $p < 0.0001$). The buying condition again showed greater sensitivity (58.6% versus 78.6%) than the selling condition (62.3% versus 68.1%, $F(1, 268) = 8.8$, $p < 0.01$).

Sensitivity of Strength Ratings to Class Factors. Log-odds-transformed strength ratings for 10 companies were regressed onto the associated aggregate cue values. Strength ratings were strongly related to the case-based cues (average slope = 0.51, ranging from 0.44 to 0.55 across different conditions). There were again no significant differences in the sensitivity of strength ratings across class characteristics or task type (setting buying or selling prices). The strength ratings can again be characterized as fully case based, consistent with H1A.

Sensitivity of Price Judgments to Class Factors. Judgment models predicting transformed buying or selling prices from the aggregate cue values were fit for each participant, and the average estimated parameters are displayed in Figure 7. As before, there was some sensitivity to outcome base rate,

Figure 7 Average Judgment Models and Bayesian Standard for Buying and Selling Price Tasks, Study 3



with higher intercepts in the high base-rate condition ($M = -0.24$) than in the low base-rate condition ($M = -0.59$, $F(1, 264) = 9.8$, $p < 0.01$). For comparison, Bayesian judgment would entail a much larger change in intercepts, from 0.53 to -0.72 . Consistent with loss aversion, intercepts were also systematically lower in the buy condition ($M = -0.63$) than in the sell condition ($M = -0.21$, $F(1, 264) = 14.0$, $p < 0.001$). There were no interactions between class factors and buying/selling on the intercept in the judgment model.

Examining the slopes of the judgment model revealed that there was essentially no sensitivity to differences in diagnosticity. The average slope of the judgment model did not differ across diagnosticity condition ($M_{lo} = 0.605$ and $M_{hi} = 0.618$), and this insensitivity to diagnosticity was similar for both buying and selling, $F_s < 1$. However, consistent with H5B, buying prices (average slope = 0.71) were substantially more sensitive to case-based cues than were selling prices (average slope = 0.52, $F(1, 264) = 24.5$, $p < 0.001$).

Both buying and selling prices were largely case based and generally insensitive to class characteristics, with insufficient sensitivity to base rate (H2B) and no sensitivity to diagnosticity (H2A). Notably, although participants in the buying condition, compared to those in the selling condition, gave direct estimates of base rate and diagnosticity that were more sensitive to the experimental manipulations, the buying prices they set based on case-specific evidence were no more sensitive than selling prices to these class-based considerations. Any greater attentional focus or accuracy motivation that may have been engendered by the buying task did not appear to manifest itself in the prices assigned to individual certificates.

Following the logic introduced in Study 1, we can attempt to reconstruct the probability weighting func-

tion for the buying and selling conditions. For each bin of cue values, we calculate the average probability judgment (from the probability condition in Study 1) and the average set prices (from the buying and selling conditions in Study 3) using only the high base-rate condition, which was unchanged across studies. The top panel of Figure 8 shows the usual inverse S-shaped function for the selling prices. The bottom panel shows the shape of the corresponding function for the buying prices, which is notably steeper and less S-shaped, as predicted. This is because the buying price plot combines two effects: the probability weighting function, which blunts the relation between case-based evidence and buying prices, and the transition across the reference point separating gains and losses, which sharpens the relation between case-based evidence and buying prices.

Summary of Empirical Results

Table 2 summarizes the tests of the various hypotheses over the studies.

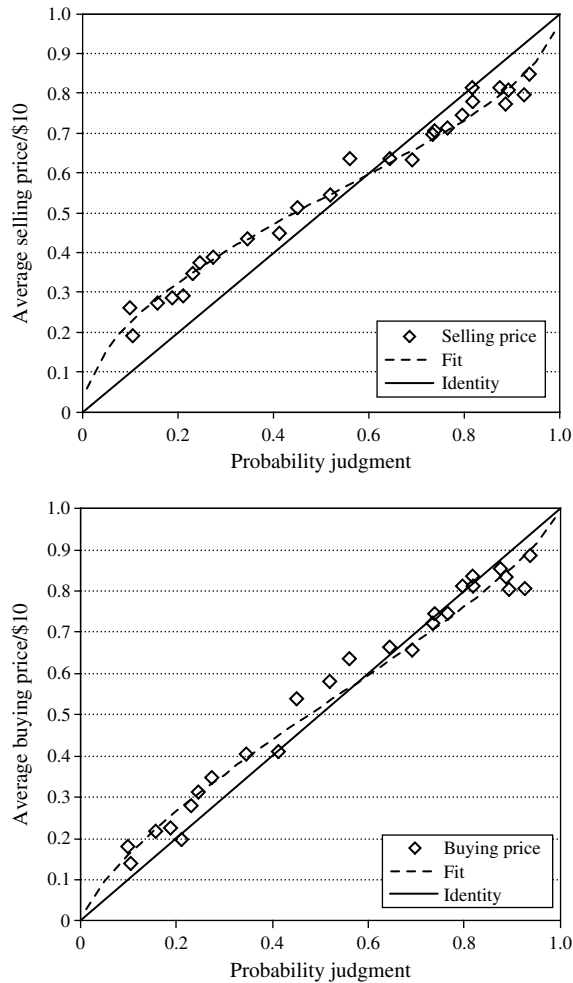
System 1 Sensitivity

System 1 impressions were sensitive to appropriate factors. Strength ratings for individual companies were sensitive to case-specific evidence and insensitive to class factors (H1A: fully case based). System 1 evaluations of the aggregate environment were consistently sensitive to base rate, but sensitive to diagnosticity differences only with very extensive experience or when setting buying prices.

Class Sensitivity of Prices

The overall characterization of prices attached to individual cases is largely consistent with case-based judgment, with some (insufficient) sensitivity to base rate (H2B) but no sensitivity to diagnosticity (H2A). As noted above and consistent with H1A, in all cases where direct strength ratings were measured,

Figure 8 Relationship Between Study 3 Buying/Selling Prices and Study 1 Judged Probabilities



Notes. Average selling prices versus average judged probability (top panel). Average buying prices versus average judged probability (bottom panel).

they showed no sensitivity to class factors. This null effect on the strength ratings suggests that any observed sensitivity to base rate is driven by System 2 adjustments rather than incorporated into a System 1 impression (in which case base-rate sensitivity would be seen in the strength ratings as well).

Economic Impact

There was no evidence of improved class sensitivity for prices compared to probability judgments (H3, H4). Differences in case sensitivity for probabilities, buying prices, and selling prices were consistent with psychophysical transformations of the prospect theory value and weighting functions (H5A, H5B). Judgments in the asset pricing task are thus predictably different than direct judgments of probability but do not exhibit greater class sensitivity, via either System 1 or System 2 processes.

As noted earlier, and illustrated schematically in Figure 3, case-based judgment implies predictable patterns of miscalibration. Figure 9 shows the aggregate calibration curves across all of the studies (including the replication of Study 1) for the different tasks: probability judgment, selling prices, and buying prices. Note that the qualitative pattern for all three tasks is similar to the case-based pattern depicted in Figure 3. Changes in base rate affect the elevation, and changes in diagnosticity affect the slope of the curves. The calibration curves illustrate the costs of insensitivity to class characteristics. In general, judgments are too high when base rate is low, and too low when base rate is high. Judgments are in general too extreme, and especially extreme when diagnosticity is low.

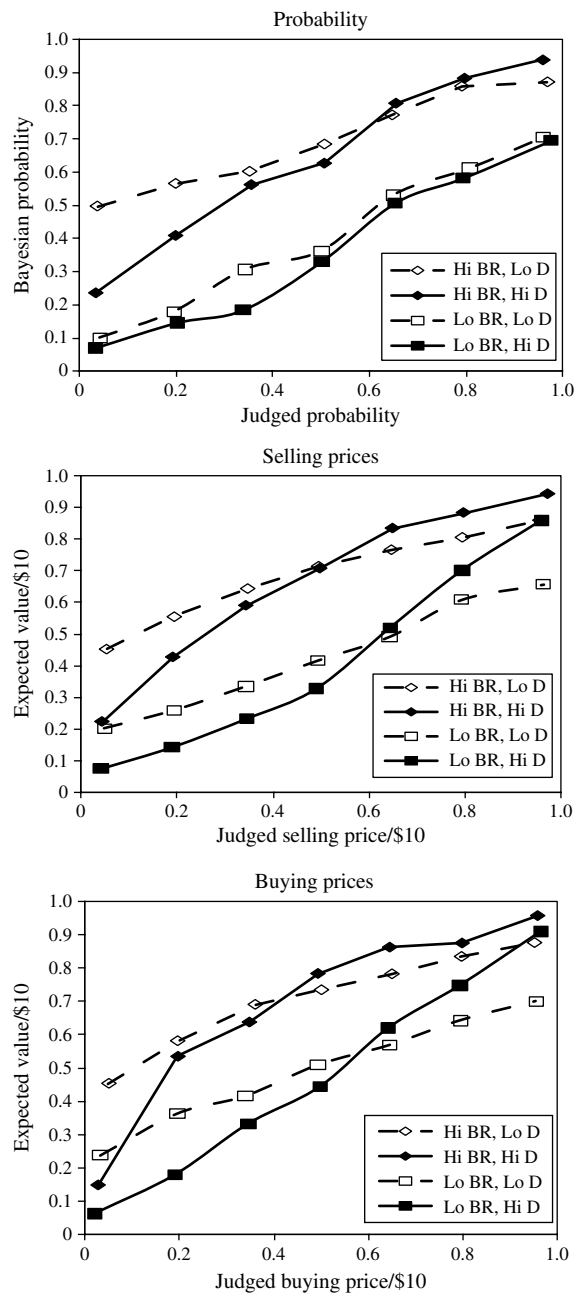
Notably, the lone case of quite good calibration is for buying prices under high diagnosticity and low base rate. Case-based judgment implies that good calibration is possible when class conditions happen to match the case-based judgment model that the judge is using. Finding good calibration in some particular circumstances is a weak test of judgmental quality. To truly examine the correspondence of judgment to external accuracy criteria, it is essential to study judgment under varying class conditions (i.e., probabilistic environments).

We fit an overall case-based model to the data, using the median judgment model intercept and slope across all subjects, irrespective of class condition. The fit of this fully case-based model to the individual data is contrasted with the Bayesian model fit to the

Table 2 Summary of Results

	Study 1	Study 2	Study 3
H1A (Strength ratings System 1 specialization; sensitive to case, insensitive to class)	Yes	—	Yes
H1B (Sensitivity of aggregate judgments to class factors)	Yes for base rate No for diagnosticity	Yes for base rate Yes for diagnosticity	Yes for base rate Yes for diagnosticity
H2A (No class sensitivity)	Yes for diagnosticity	Yes for diagnosticity	Yes for diagnosticity
H2B (Insufficient class sensitivity)	Yes for base rate	Yes for base rate	Yes for base rate
H2C (Bayesian class sensitivity)	No	No	No
H3 (Pricing more sensitive to class than probability via System 1)	No	—	—
H4 (Pricing more sensitive to class than probability via System 2 correction)	No	—	—
H5A (Selling prices less sensitive to case evidence than probability judgments)	Yes	—	—
H5B (Selling prices less sensitive to case evidence than buying prices)	—	—	Yes

Figure 9 Calibration Curves Across All Studies



observed average judgment for each condition and task. The average absolute deviation is 13% for the Bayesian model (12.7% for probability, 12.9% for selling prices, and 13.7% for buying prices) but only 7% for the case-based model (5.8% for probability, 7.3% for selling prices, and 7.8% for buying prices). In terms of simple descriptive characterizations, case-based judgment is distinctly superior to Bayesian judgment.

General Discussion

We investigated whether case-based patterns of miscalibration typically observed in probability judgments

would be reduced or eliminated in a more economically relevant task such as setting prices for assets. Across a series of studies using a simulated stock market where participants actively learned about the value of evidence and the overall bearishness or bullishness of the market environment, we found no support for the economic impact hypothesis. Instead, case-based judgment with its signature insensitivity to class-based factors was characteristic of price setting as well as probability judgment. However, other novel predictions derived from a two-systems case-based model were confirmed: selling prices were less sensitive to case-specific cues than were probability judgments, and buying prices were more sensitive to case-specific cues than were selling prices. Unexpectedly, compared to setting selling prices, setting buying prices made respondents more sensitive to class factors in their aggregate judgments of the information environment, although they were no more likely to use the relevant class considerations in making their individual price assignments.

These findings have a number of implications for the study and understanding of behavioral finance. First, the characteristic biases of case-based judgment, as described by the heuristics and biases program (e.g., Kahneman and Tversky 1973, Tversky and Kahneman 1974) and its extensions (the strength-weight theory of Griffin and Tversky 1992 and support theory of Tversky and Koehler 1994), are robust when respondents set prices in an experimental prediction market setting and are not diminished by extensive economic feedback and experience in that setting. This implies that a fruitful direction for identifying anomalies would be in classifying the evidential weight of different predictive cues (e.g., earnings announcements, weather forecasts, bond ratings, sports betting tips) and looking for specific patterns of overreaction and underreaction.

Second, the robust pattern of case-based miscalibration found in both pricing and probability judgment is a reminder that biases in uncertainty assessment are not limited to overconfidence and optimism, the most common sources of psychologically grounded hypotheses in behavioral finance and economics. The case-based model implies diverse patterns of calibration that depend crucially on the particular class features of the judgment environment. Understanding the information environment is at least as important as understanding the types of decision makers in predicting the type of bias to be found.

Third, the effects of loss aversion, one of the most important tools in the behavioral finance toolbox, also remain robust even when participants make rapid trial by trial pricing judgments. This speaks to a current controversy about whether ownership of an object is required to induce loss aversion on selling

versus buying prices (Morewedge et al. 2009). More generally, the results draw attention to the broad applicability and relevance of the two psychophysical functions described by prospect theory. Related to this, the reduced sensitivity to case-based cues found in pricing compared to probability judgment is a reminder that moving to an economic setting does not sharpen judgment at every level: in fact, in some cases the economic setting led to larger biases (see Figure 4).

Perhaps most notable is the difficulty participants had in encoding and applying the predictive value of cues. Participants were able to accurately encode differences in base rate of success and at least partially incorporate the base rate into their prices. However, differences in cue validity between diagnosticity conditions consistently failed to register in overall judgments or in trial by trial pricing or probability judgments. This insensitivity to diagnosticity caused substantial differences in calibration accuracy across conditions (as illustrated in the aggregate calibration curves in Figure 9), in which prices set under weakly diagnostic cues tended to be much too extreme. To the extent that most realistic judgment environments are characterized by generally weak cues, case-based judgment that is insensitive to evidence quality will often produce overextremity. A very strong-looking asset will be overvalued, and a weak-looking asset will be undervalued.

All tasks used in these studies examined judgments made by independent individuals in an incentive-compatible simulated marketplace; the question naturally arises as to whether these biases would be reduced in an interactive market setting or with domain experts. In ongoing research, we have examined whether experimental trading markets further moderate these effects: the general pattern of biases remain, with complete diagnosticity neglect but reduced base-rate neglect compared to the individual judgment. We hope that the models and data presented here will help to provide additional pathways to link laboratory research in judgment and decision making with field research in behavioral finance.

Acknowledgments

Financial support was provided by the Warrington College of Business (Brenner), the Social Sciences and Humanities Research Council of Canada (Griffin), and the Natural Sciences and Engineering Research Council of Canada (Koehler).

Appendix A. Derivation of Bayesian Slope and Intercept

Consider the values of the judgment model parameters α and β consistent with perfect (Bayesian) calibration. First define the two conditional cue distributions, for success trials and for failure trials. Let the distribution of aggregate cues C for increase trials (successes) be Gaussian with a mean of μ_1 and variance σ^2 and the distribution of C for

decrease trials (failures) be Gaussian with mean of μ_0 and the same variance σ^2 . Define μ as the mean of μ_0 and μ_1 and $d = (\mu_1 - \mu_0)/\sigma$ is the difference between the means of these two distributions measured in standard units. B represents the base rate of success trials.

Based on the Gaussian density function, the log-likelihood ratio of the case evidence (for increase relative to decrease trials) is

$$\ln\left(\frac{f(C|Inc)}{f(C|Dec)}\right) = -\frac{(\mu_1^2 - \mu_0^2)}{2\sigma^2} + \frac{\mu_1 - \mu_0}{\sigma^2}C.$$

Inserting into Bayes' rule in log-odds form,

$$\ln\left(\frac{P(Inc|C)}{P(Dec|C)}\right) = \ln\left(\frac{P(Inc)}{P(Dec)}\right) + \ln\left(\frac{f(C|Inc)}{f(C|Dec)}\right).$$

The result is the Bayesian log-odds of success, given a particular cue value C :

$$\ln\left(\frac{P(Inc|C)}{P(Dec|C)}\right) = \left[\ln\left(\frac{P(Inc)}{P(Dec)}\right) - \frac{(\mu_1^2 - \mu_0^2)}{2\sigma^2}\right] + \left(\frac{\mu_1 - \mu_0}{\sigma^2}\right)C.$$

This expression for the Bayesian log-odds is a linear function of C , as is the model of judgment. Therefore, the ideal parameters for the judgment model (to achieve optimal calibration) match the intercept and the slope. The optimal intercept and slope are then

$$\alpha^* = \left[\ln\left(\frac{P(Inc)}{P(Dec)}\right) - \frac{(\mu_1^2 - \mu_0^2)}{2\sigma^2}\right] = \ln\left(\frac{B}{1-B}\right) - \frac{d\mu}{\sigma},$$

$$\beta^* = \frac{\mu_1 - \mu_0}{\sigma^2} = \frac{d}{\sigma}.$$

Appendix B. Model for Buying Prices (Study 3)

Consider setting a maximum buying price B for an asset that pays \$1 for a success (stock price increase) and \$0 for a failure. The asset has case-based evidence C associated with it.

The subjective probability P is again modeled as a function of the case-based evidence C ; specifically the logit-transformed P is a linear function of C : $\ln(P/(1-P)) = L_P = \alpha + \beta C$. The buying price B is set so that the judge is indifferent between acquiring the certificate at price B and the status quo of not buying the asset. Acquiring the asset while paying B leads to possible payoffs of $1 - B$ under the success outcome (with subjective probability P) and $-B$ under the failure outcome. Therefore, the buying price B must satisfy

$$0 = w(P)v(1-B) + (1-w(P))v(-B).$$

Rearranging, we get

$$\frac{-v(-B)}{v(1-B)} = \frac{w(P)}{1-w(P)}.$$

We consider a traditional power value function

$$v(x) = x^{1/\theta} \quad \text{for } x \geq 0,$$

$$v(x) = -\lambda(-x)^{1/\theta} \quad \text{for } x < 0.$$

The exponent $1/\theta$ indexes the curvature of the function, with risk aversion for gains with $\theta > 1$. The coefficient λ indexes loss aversion, the extent to which losses loom larger than gains.

Again we use the Gonzalez and Wu (1999) “linear in log-odds” weighting function:

$$w(P) = \frac{\delta P^\gamma}{\delta P^\gamma + (1-P)^\gamma}.$$

Inserting the weighting and value functions, we get

$$\lambda \left(\frac{B}{1-B} \right)^{1/\theta} = \delta \left(\frac{P}{1-P} \right)^\gamma.$$

This equation then simplifies to a linear function in log-odds:

$$\ln \frac{B}{1-B} = L_B = \theta \ln \delta - \theta \ln \lambda + \theta \gamma L_P.$$

Because L_P is a linear function of the case-based evidence C , L_B is also

$$L_B = (\theta \ln \delta - \theta \ln \lambda + \theta \gamma \alpha) + (\theta \gamma \beta) C.$$

Compare the buying-price model to the selling-price model:

$$L_S = (\ln \delta - \ln \theta) + (\gamma \beta) C.$$

Note that the slope for the buying price model contains the value function parameter θ , which is expected to be greater than one to entail risk aversion for gains and risk seeking for losses in the S-shaped prospect theory value function. Based on this, we predict shallower slopes for selling prices than for buying prices (H5B). The comparison of the intercepts does not yield a straightforward prediction.

References

- Barberis, N., A. Shleifer, R. Vishny. 1998. A model of investor sentiment. *J. Financial Econ.* 49(3) 307–343.
- Becker, G. M., M. H. DeGroot, J. Marschak. 1964. Measuring utility by a single-response sequential method. *Behav. Sci.* 9(3) 226–232.
- Bloomfield, R. 1996. Quotes, prices, and estimates in a laboratory market. *J. Finance* 51(5) 1791–1808.
- Brenner, L., D. Griffin, D. Koehler. 2005. Modeling patterns of probability calibration with random support theory: Diagnosing case-based judgment. *Organ. Behav. Human Decision Processes* 97(1) 64–81.
- Brenner, L., D. J. Koehler, Y. Rottenstreich. 2002. Remarks on support theory: Recent advances and future directions. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, UK, 489–505.
- Camerer, C. 1992. The rationality of prices and volume in experimental markets. *Organ. Behav. Human Decision Processes* 51(2) 237–272.
- Daniel, K. D., D. Hirshleifer, A. Subrahmanyam. 2001. Overconfidence, arbitrage, and equilibrium asset pricing. *J. Finance* 56(3) 921–965.
- Evans, J. St. B. T. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Rev. Psych.* 59 255–278.
- Fox, C. R. 1999. Strength of evidence, judged probability, and choice under uncertainty. *Cognitive Psych.* 38(1) 167–189.
- Fox, C. R., A. Tversky. 1998. A belief-based account of decision under uncertainty. *Management Sci.* 44(7) 879–895.
- Fox, C. R., B. A. Rogers, A. Tversky. 1996. Options traders exhibit subadditive decision weights. *J. Risk Uncertainty* 13(1) 5–19.
- Gilovich, T. D., D. Griffin. 2002. Introduction—Heuristics and biases: Then and now. D. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, New York, 1–18.
- Gonzalez, R., G. Wu. 1999. On the shape of the probability weighting function. *Cognitive Psych.* 38(1) 129–166.
- Griffin, D., L. Brenner. 2004. Perspectives on probability judgment calibration. D. J. Koehler, N. Harvey, eds. *Blackwell Handbook of Judgment and Decision Making*. Blackwell, Oxford, UK, 177–199.
- Griffin, D., A. Tversky. 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psych.* 24(3) 411–435.
- Hsee, C. K., Y. Rottenstreich. 2004. Music, pandas, and muggers: On the affective psychology of value. *J. Experimental Psych.: General* 133(1) 23–30.
- Hsu, M., I. Krajbich, C. Zhao, C. F. Camerer. 2009. Neural response to reward anticipation under risk is nonlinear in probabilities. *J. Neuroscience* 29(7) 2231–2237.
- Kahneman, D., S. Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, UK, 49–81.
- Kahneman, D., A. Tversky. 1973. On the psychology of prediction. *Psych. Rev.* 80(4) 237–251.
- Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2) 263–292.
- Kahneman, D., J. Knetsch, R. Thaler. 1991. The endowment effect, loss aversion, and the status quo bias. *J. Econom. Perspect.* 5(1) 193–206.
- Kahneman, D., P. Slovic, A. Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York.
- Kirchler, E., B. Maciejovsky. 2002. Simultaneous over- and underconfidence: Evidence from experimental asset markets. *J. Risk Uncertainty* 25(1) 65–85.
- Koehler, D. J. 1996. A strength model of probability judgments for tournaments. *Organ. Behav. Human Decision Processes* 66(1) 16–21.
- Koehler, D. J., L. Brenner, D. Griffin. 2002. The calibration of expert judgment: Heuristics and biases beyond the laboratory. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, UK, 686–715.
- Koehler, D. J., L. A. Brenner, A. Tversky. 1997. The enhancement effect in probability judgment. *J. Behav. Decision Making* 10(4) 293–313.
- Massey, C., G. Wu. 2005. Detecting regime shifts: The causes of under- and overreaction. *Management Sci.* 51(6) 932–947.
- Morewedge, C. K., L. L. Shu, D. T. Gilbert, T. D. Wilson. 2009. Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *J. Experimental Soc. Psych.* 45(4) 947–951.
- Odean, T. 1999. Do investors trade too much? *Amer. Econom. Rev.* 89(5) 1279–1298.
- Rick, S. 2010. Losses, gains, and brains: Neuroeconomics can help answer open questions about loss aversion. *J. Consumer Psych.* 21(4) 453–463.
- Rottenstreich, Y., C. K. Hsee. 2001. Money, kisses, and electric shocks: On the affective psychology of risk. *Psych. Sci.* 12(3) 185–190.
- Sloman, S. A. 1996. The empirical case for two systems of reasoning. *Psych. Bull.* 119(1) 3–22.
- Stanovich, K. E., R. F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behav. Brain Sci.* 23(5) 645–665.
- Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185(4157) 1124–1131.
- Tversky, A., D. Kahneman. 1991. Loss aversion in riskless choice: A reference dependent model. *Quart. J. Econom.* 106(4) 1039–1061.
- Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4) 297–323.
- Tversky, A., D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psych. Rev.* 101(4) 547–567.