A Behavioral Demonstration of Overconfidence in Judgment
Albert E. Mannes[a]
Don A. Moore[b]

[a]University of Pennsylvania, 517 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104.
[b]University of California at Berkeley, 2222 Piedmont Avenue, Berkeley, CA 94720.

Abstract

Overprecision—an excessive faith that one knows the truth—is both the most durable and least understood form of overconfidence. This paper offers an approach to the study of overprecision that has more in common with the way uncertainty affects our actions in everyday life and avoids some methodological problems of other approaches. We measure the precision in judgment implied by people's tendency to adjust their point estimates of an uncertain quantity in response to the costs of over- or underestimating the correct answer. The results reveal robust overprecision. People adjust their estimates less than they should given their actual knowledge, and this is driven by their subjective confidence.

Author note: Data and stimuli are available at http://learnmoore.org/mooredata/EJM/. Address correspondence to albert.mannes@gmail.com or don.moore@alumni.carleton.edu.

A Behavioral Demonstration of Overconfidence in Judgment

Overconfidence has been called the most pervasive and potent bias to which human

judgment is vulnerable (De Bondt & Thaler, 1995; Kahneman, 2011). Of the various ways in

which overconfidence has been studied, the most robust and least understood is what (Moore &

Healy, 2008) call overprecision—excessive confidence that one knows the truth. Overprecision

has profound consequences, giving investors excessive confidence in how much an investment is

worth (Daniel, Hirshleifer, & Subrahmanyam, 2001), leading physicians to gravitate too quickly

to a particular diagnosis (Christensen-Szalanski & Bushyhead, 1981), and making us intolerant

of dissenting views (Harvey, 1997; Harvey & Fischer, 1997).

Using standard approaches, overprecision has proven remarkably resistant to debiasing

(Harvey, 1997; Soll & Klayman, 2004). This has led to concerns that overprecision may be a

methodological artifact (Dawes & Mulford, 1996; Gigerenzer, 1991). Researchers typically

study overprecision by asking people to specify 90% confidence intervals around some

numerical estimate (e.g., the length of the Nile River). These confidence intervals include the

truth as rarely as 30% of the time, suggesting people are behaving as if their knowledge was

more precise than it actually is (McKenzie, Liersch, & Yaniv, 2008; Soll & Klayman, 2004).[1]

One criticism of this approach is that it requires familiarity with probability and confidence

intervals—statistical concepts with which even well-educated people routinely make large errors

(Cosmides & Tooby, 1996; Tversky & Kahneman, 1993).

A second criticism is that this approach bears little relationship to the way overprecision

affects people's judgments in daily life. It is rare for people to have to specify confidence

intervals around some belief. When deciding what time to depart for a lunch meeting, one does

not consider a 90% confidence interval around the travel time. Instead, if punctuality is

important, then uncertainty should lead one to depart earlier. Or because falling off a cliff is worse than missing the great view, we keep back from the precipice. Every day, in hundreds of different instances, we steer clear of metaphorical cliffs as a function of (1) the cost of erring on each side and (2) our uncertainty. In this paper we manipulate these two factors to observe their effects on people's judgments. The first benefit of our approach is that it does not require us to ask participants about probabilities, degrees of confidence, or confidence intervals. Second, our approach parallels the way in which uncertainty affects judgment every day.

Our approach builds on that of Mamassian (2008) to test overprecision in declarative knowledge. In our studies, people estimated the high temperature of cities on randomly selected days. We gave them feedback on the accuracy of their estimates and rewarded them for accuracy with lottery tickets toward a prize. The days were divided into three blocks of 24 judgments each, and each block featured a different payoff function (see Table 1). One block featured symmetric payoffs, in which participants earned lottery tickets if their estimates were close to the correct answer. The two other blocks featured asymmetric payoffs. In one, participants earned lottery tickets only for correctly guessing or overestimating the answer; in the other, participants earned lottery tickets only for correctly guessing or underestimating the answer. These payoffs were designed to mimic the asymmetric costs of erring to one side encountered in daily life.

To illustrate, imagine a person whose best guess of the high temperature in Pittsburgh on March 10, 2006 is 10°C, but who also believes it could have ranged from 7°C to 12°C. Under symmetric payoffs, in which the costs of over- and underestimation are identical, she should guess 10°C. But because 7°C is also reasonable, she should increase her guess to, say, 13°C (an adjustment of +3°C) if she is rewarded only for correctly guessing or overestimating the actual

temperature. At the same time, however, she will only earn two tickets instead of five if she overestimates by too much (see Table 1, Block 2). Thus to maximize earnings, her guess should be close to the correct answer but not less. Note that confidence in one's knowledge should affect the degree to which one adjusts the estimates under asymmetric payoffs. If, for example, a less confident person believes the temperature could have ranged from 5°C to 15°C, he should make larger adjustments to his estimates (e.g., +5°C) to maximize his expected earnings.

We make two predictions about people's behavior in this task. First, we expect people to be overprecise: They will adjust their point estimates insufficiently. Experiments 1A and 1B demonstrate this basic result. Second, we expect people's confidence to moderate their overprecision. Specifically, the more confidence people have in their point estimates, the less they will adjust. We manipulate confidence in Experiment 2.

Experiment 1A

Method

We recruited 36 women and 28 men from Pittsburgh, Pennsylvania (M_{age} = 25.6 years) to participate in a study of judgment conducted at a local university. Participants were offered $5 and lottery tickets towards an iPod Shuffle for completing the study. Participants first provided 90% confidence intervals (90% CIs) for the high temperature in Pittsburgh (in °C or °F, at their choosing) on 24 randomly-selected days in 2006–2007. This enabled us to compare traditional measures of overprecision with ours. Following this, participants made point estimates of the high temperature for 72 days, divided into three blocks of 24, and received trial-by-trial feedback with the correct answers. Participants earned tickets for the accuracy of their estimates based on the payoffs in each block, up to a maximum of 120 per block. Table 1 explains the payoffs. The block with symmetric payoffs came first, followed by the two blocks with asymmetric payoffs,

which were counterbalanced. Participants then answered questions about their knowledge of Pittsburgh temperatures and their demographic characteristics, learned how they did, and were dismissed.

Measures

We operationalized *overprecision* in the point estimates as the ratio between the observed adjustment of each person's estimates under asymmetric payoffs and a normative adjustment which maximized the expected payoffs in those blocks (cf. Granger, 1969). The observed adjustment for each person was his or her mean error (i.e., bias) in each block of estimates. We expected people to bias their estimates upward when rewarded for overestimating the correct answer (i.e., a positive average adjustment in Block 2) and to bias their estimates downward when rewarded for underestimating the correct answer (i.e., a negative average adjustment in Block 3). To identify the normative adjustment for each person, we asked a simple question: How many more tickets would the participant have earned if he or she had added another *X* degrees to the estimates, on average? We calculated these amounts for 201 values of *X* (viz., –50°C, –49.5°C, …, +49.5°C, +50°C) in each of the asymmetric-payoff blocks. We then averaged those values of *X* at which the simulated earnings exceeded the participant's actual earnings. This represented his or her normative adjustment (cf. Lawrence & O'Connor, 2005).[2]

Overprecision was indexed by the ratio of each person's observed adjustment to the normative adjustment, subtracted from 1. If the observed adjustment was less than the normative adjustment, overprecision was greater than zero; if the observed adjustment was greater than the normative adjustment, overprecision was less than zero—the person was underprecise; and if the observed and normative adjustments were equal, overprecision equaled zero—the person was neither under- nor overprecise. For example, if a person's observed adjustment is +3°C, but he

or she would have earned more with an average adjustment of +5°C (see Figure 1), the degree of overprecision is .40 (= 1 – 3/5).

Results and Discussion

Preliminary analysis. Two point estimates exceeding 100°C were coded as entry errors and excluded prior to the analysis. To check the manipulation of payoffs, we calculated the observed adjustment of each person by block. Average adjustments (in °C) were 0.69, 95% CI [0.25, 1.13], with symmetric payoffs; 4.93, 95% CI [4.35, 5.50], with positive payoffs for overestimation; and –6.29, 95% CI [–6.88, –5.70], with positive payoffs for underestimation. People adjusted (i.e., biased) their estimates as expected with this manipulation of payoffs.

Point estimates. Because differences between the two blocks with asymmetric payoffs were not of substantive interest, we averaged the observed and normative adjustments (reverse-coded as necessary) across blocks for each person before calculating overprecision. The average level of overprecision was 0.33 ($SD = 0.26$), $t(63) = 10.16$, $p < .001$, Cohen's $d = 1.27$. Participants' actual earnings under asymmetric payoffs ($M = 118.69$, $SD = 12.06$) were significantly less than what they would have earned absent their overprecision ($M = 125.05$, $SD = 8.99$), $t(63) = 19.49$, $p < .001$, $d = 0.60$.

Participants who made larger absolute errors under symmetric payoffs tended to make larger adjustments to their point estimates under asymmetric payoffs ($r_s = .28$, $p = .024$). In other words, those less knowledgeable about Pittsburgh temperatures adjusted more, as they should have. This suggests that confidence—knowledge of what we know (Russo & Schoemaker, 1992)—was an important predictor of behavior.

Confidence intervals. On average, participants' 90% CIs contained the correct answer 49% ($SD = 24\%$) of the time, which replicates prior findings using this method. There was also

correspondence in overprecision across the two methods. Fifty-four of the 64 participants were overprecise in both their confidence intervals and point estimates, seven were overprecise only in their confidence intervals, and three were overprecise only in their point estimates. Individuals making the smallest adjustments to their point estimates constructed the narrowest 90% CIs ($r_s =$ .43, $p < .001$).

Experiment 1B

Presenting the block with symmetric payoffs first in Experiment 1A provided people with extensive feedback about their actual knowledge, which enabled them to make informed judgments about how to adjust their estimates under asymmetric payoffs. However, it is possible that people found it difficult to switch from making their best estimate of the temperature under symmetric payoffs to making biased estimates under asymmetric payoffs. If so, this would overstate the extent of overprecision. With this concern in mind, Experiment 1B counterbalanced the order of the symmetric and asymmetric payoffs.

Method

We recruited 60 women and 42 men from Berkeley, California ($M_{age} = 20.7$ years) to participate. Each was offered $10 and a chance to win $50 gift cards from an online retailer. Participants guessed the high temperature in Berkeley (in °C or °F, at their choosing) on randomly chosen days from 2006–2007. The procedure and measures were identical to that of the prior experiment with two exceptions. First, we counterbalanced the order of the symmetric and asymmetric payoffs. Second, participants completed the 90% CIs after their point estimates.

Results and Discussion

Average adjustments (in °C) were –1.09, 95% CI [–1.34, –0.84], with symmetric payoffs; 2.79, 95% CI [2.46, 3.12], with positive payoffs for overestimation; and –2.84, 95% CI [–3.12, –2.57], with positive payoffs for underestimation.

The average level of overprecision was 0.40 ($SD = 0.26$), $t(101) = 15.82$, $p < .001$, $d = 1.57$. Overprecision was marginally higher for the 53 participants who started with the asymmetric payoffs ($M = 0.45$, $SD = 0.26$) than for the 49 participants who started with the symmetric payoffs ($M = 0.36$, $SD = 0.25$), $t(100) = 1.70$, $p = .092$, $d = 0.34$. This suggests that presenting the symmetric payoffs first, if anything, produced less overprecision.

Confidence intervals. On average, participants' 90% CIs contained the correct answer 70% ($SD = 18\%$) of the time. As in the prior study, individuals making the smallest adjustments to their point estimates constructed the narrowest 90% CIs ($r_s = .38$, $p < .001$). Of the 102 participants, 75 were overprecise in both their confidence intervals and point estimates, eight were overprecise only in their confidence intervals, 18 were overprecise only in their point estimates, and one was overprecise in neither task.

Experiment 2

Experiments 1A and 1B produced strong evidence of overprecision in judgments. Experiment 2 demonstrates that confidence moderates people's point-estimate adjustments. We manipulated confidence directly by varying the feedback participants received (cf. Arkes, Christensen, Lai, & Blumer, 1987; Stone & Opel, 2000). Specifically, participants received no trial-by-trial feedback about their accuracy, authentic feedback (as in the prior experiments), or exaggerated feedback. For those in the no-feedback and authentic-feedback conditions, we expected to replicate the overprecision found in the first experiments. We expected those

receiving exaggerated feedback to show the least overprecision.

Method

We recruited 107 women and 77 men from Berkeley, California (M_{age} = 20.5 years) to participate. Each was offered $10 and a chance to win $25 gift cards from an online retailer. The procedure was the same as that of Experiment 1A with two exceptions. First, after specifying their 90% CIs, participants were assigned to one of three conditions for the point estimates. In the *no-feedback* condition, participants completed their point estimates without trial-level or summary feedback about their errors; in the *authentic-feedback* condition, participants received accurate trial-level feedback about their errors; and in the *exaggerated-feedback* condition, participants received manufactured trial-level feedback about their errors. The exaggerated feedback told participants their errors were approximately 2.5 times greater than they actually were (e.g., if the actual error was +2°C, the feedback was +5°C). We selected this exaggeration after estimating what feedback might have optimized accuracy in Experiments 1A and 1B.

Second, after making their point estimates under symmetric payoffs, all participants provided subjective probability distributions of their errors based on the SPIES technique (Haran, Moore, & Morewedge, 2010). They indicated, for every 100 temperature estimates, how many of their errors would fall into each of nine intervals (viz., < −11°C, −10 to −8°C,…, +8 to +10°C, > +11°C). After assigning frequencies to each interval, participants completed their point estimates under asymmetric payoffs. Participants then answered questions about their understanding of the task, their motivation, and their confidence, after which they learned their performance and were dismissed.

Measures

Overprecision was operationalized as in the prior experiments. We collected two measures of participants' confidence. The implicit measure was the subjective probability distribution of their errors provided after the first block of point estimates. We expected participants with less confidence to construct wider probability distributions (as measured by the standard deviation of their distributions) than those with more confidence (Haran, et al., 2010). The explicit measure of confidence was each person's answer to the following question collected at the end of the study: "How confident are you in your ability to accurately estimate Berkeley, CA temperatures?" (0 = *Not at all confident*, 4 = *Extremely confident*).

Results and Discussion

Preliminary analysis. Four estimates exceeding 100°C in absolute value were dropped as entry errors prior to analysis. In addition, three participants who made constant and extreme estimates (for Berkeley, CA) under asymmetric payoffs (e.g., –18°C) were excluded from the analysis, leaving a final sample of 181.[3]

Average adjustments (in °C) were –0.29, 95% CI [–0.47, –0.10], with symmetric payoffs; 3.78, 95% CI [3.40, 4.15], with positive payoffs for overestimation; and –3.21, 95% CI [–3.63, –2.80], with positive payoffs for underestimation.

Confidence. The correlation between the two measures of confidence was –.41 (i.e., people with wider probability distributions also reported lower confidence). A multivariate omnibus test rejected that feedback had no effect on confidence, $F(4, 178) = 13.20$, $p < .001$. As Table 2 indicates, for both the subjective probability distributions, $F(2, 178) = 16.33$, $p < .001$, $\eta^2 = .15$, and self-reported confidence, $F(2, 178) = 18.28$, $p < .001$, $\eta^2 = .17$, participants in the no-feedback condition were most confident (i.e., had the narrowest probability distributions),

followed by those in the authentic-feedback condition, followed by those in the exaggerated-feedback condition.

Point estimates. The average level of overprecision was .29 (SD = .48), $t(180)$ = 8.31, p < .001, d = 0.62. Participants' total earnings under asymmetric payoffs (M = 116.40, SD = 18.77) were significantly less than what they would have earned absent their overprecision (M = 128.86, SD = 12.35), $t(180)$ = 17.81, p < .001, d = 0.79.

As expected, we reduced overprecision by manipulating confidence. Table 2 indicates that participants receiving exaggerated feedback were significantly less overprecise than those receiving either no feedback, $t(178)$ = 3.91, p < .001, d = 0.59, or authentic feedback, $t(178)$ = 4.95, p < .001, d = 0.74. In fact, overprecision in the exaggerated-feedback condition did not differ statistically from zero, $t(58)$ = 0.72, p = .47, d = 0.09.

People display overprecision in this task because they insufficiently adjust their point estimates relative to their normative benchmarks—they step too close to the cliff's edge. We have argued these adjustments are affected by confidence: People who are less confident in their expertise adjust more than people who are more confident. The data supported this claim. Participants with wider subjective probability distributions made larger adjustments to their subsequent point estimates (r = .37, p < .001) and were less overprecise (r = −.28, p < .001).

Confidence intervals. On average, participants' 90% CIs contained the true answer 63% (SD = 22%) of the time. This overprecision did not differ by condition, $F(2, 178)$ = 0.91, p = .405. Individuals making the smallest adjustments to their point estimates constructed the narrowest 90% CIs (r_s = .18, p = .018). Of the 181 participants, 121 were overprecise in both their confidence intervals and point estimates, 37 were overprecise only in their confidence intervals, 17 were overprecise only in their point estimates, and six were overprecise in neither

task.

Discussion

The three experiments we report present evidence of overprecision in judgment using a behavioral measure that (1) does not require an understanding of probabilities or confidence intervals by research participants and (2) bears more similarity to the everyday decision contexts that depend on confidence in judgment. The results suggest that overprecision is robust and is not solely a product of unfamiliar elicitation methods (such as confidence intervals) that assume too much about respondents' statistical sophistication.

Yet two concerns persist. First, participants may have simply been insensitive to the asymmetric payoffs. Anticipating this, we took measures to ensure participants understood the task: They had to correctly answer three questions about the payoffs to demonstrate their attention and understanding before starting each block, and the payoff function was summarized at the top of the computer screen throughout the trials. Moreover, as reported above, participants biased their estimates in the appropriate direction, and did so, on average, on the first trial of each block. In short, participants understood their incentives (for additional evidence about people's ability to respond to asymmetric payoffs, see Goodwin, 2005; Lawrence & O'Connor, 2005; Weber, 1994).

Second, even if they understood their payoffs, participants may have insufficiently adjusted their estimates because they were anchored by their best estimate of the answer (Block & Harper, 1991). In fact, verbal protocols from a supplementary study confirmed that they often started with a best guess and adjusted from there. However, while anchoring processes may have been at work, they cannot explain why (a) those with less expertise made greater adjustments in Experiment 1A, (b) those with less confidence made greater adjustments in

Experiment 2, and (c) overprecision was eliminated in Experiment 2 when people received unflattering feedback about their accuracy. In sum, when asymmetric incentives lead people to adjust their responses to one side or to the other of the best guess, the amount of adjustment depends critically on how confident people are in the accuracy of their knowledge—the subjective precision in their judgment. Excessive confidence is therefore key to participants' failure to adjust sufficiently.

The new method we present has its limitations. It does not elicit direct measures of confidence, and it is pedagogically less useful than traditional means of illustrating overconfidence, such as confidence-interval estimation (Russo & Schoemaker, 1992). Nevertheless, it offers researchers an alternative to self-reports which rely on having to explain probabilities or confidence intervals.

The consequence of all this overprecision is profound. It happens with some frequency that people cut things too close—arriving late, missing planes, bouncing checks, or falling off one of the many cliffs that present themselves to us in our daily lives. People also cling too fervently to beliefs that are poorly supported by evidence, adjusting their beliefs too little in light of the evidence or the consequences of being wrong. We hope that the methodology and the evidence we present here will prove useful in elucidating the underlying sources of this excessive faith in the accuracy of our own judgment.

References

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing

overconfidence. *Organizational Behavior and Human Decision Processes, 39*(1), 133-

144.

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-

adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49*(2),

188-207.

Bolger, F., & Harvey, N. (1995). Judging the probability that the next point in an observed time-

series will be below, or above, a given value. [Article]. *Journal of Forecasting, 14*(7),

597-607. doi: 10.1002/for.3980140705

Christensen-Szalanski, J. J., & Bushyhead, J. B. (1981). Physicians' use of probabilistic

information in a real clinical setting. *Journal of Experimental Psychology: Human

Perception and Performance, 7*(4), 928-935.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking

some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1-

73.

Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (2001). Overconfidence, arbitrage, and

equilibrium asset pricing. *Journal of Finance, 56*(3), 921-965.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in

judgment or flaws in how we study judgment? *Organizational Behavior and Human

Decision Processes, 65*(3), 201-211.

De Bondt, W. F. M., & Thaler, R. H. (1995). Financial decision-making in markets and firms: A

behavioral perspective. In R. Jarrow, V. Maksimovic & W. T. Ziemba (Eds.), *Handbooks*

in Operations Research and Management Science (Vol. 9, pp. 385-410). Amsterdam:

Elsevier.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "Heuristics and

Biases". In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology*

(Vol. 2, pp. 83-115). Chichester, UK: John Wiley and Sons, Ltd.

Goodwin, P. (2005). Providing support for decisions based on time series information under

conditions of asymmetric loss. [Article]. *European Journal of Operational Research,*

163(2), 388-402. doi: 10.1016/j.ejor.2003.10.039

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational*

Research Quarterly, 20(2), 199-207.

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in

judgment. *Judgment and Decision Making, 5*(7), 467-476.

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences, 1*(2), 78-82.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and

sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*(2),

117-133.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.

Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions.

[Article]. *International Journal of Forecasting, 21*(1), 3-14. doi:

10.1016/j.ijforecast.2004.02.003

Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science, 19*(6), 601-606. doi: 10.1111/j.1467-9280.2008.02129.x

McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes, 107*(2), 179-191. doi: 10.1016/j.obhdp.2008.02.007

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517.

Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review, 33*(2), 7-17.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299-314.

Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational Behavior and Human Decision Processes, 83*(2), 282-309.

Tversky, A., & Kahneman, D. (1993). Probabilistic reasoning. In A. I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 43-68). Cambridge, MA, USA: Mit Press.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. [Article]. *Psychological Bulletin, 115*(2), 228-242. doi: 10.1037/0033-2909.115.2.228

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. [Article]. *Journal of Experimental Psychology: Learning Memory and Cognition, 30*(6), 1167-1175. doi: 10.1037/0278-7393.30.6.1167

Footnotes

[1] Elicitation methods that call explicit attention to a large range of outcomes can reduce overprecision (see Haran, et al., 2010; Winman, Hansson, & Juslin, 2004), and, in some cases, even cause underprecision (Bolger & Harvey, 1995; Goodwin, 2005).

[2] We relied on each person's empirical distribution of errors for each payoff function to identify his or her normative adjustment rather than make parametric assumptions about those distributions, such as normality or invariance across payoff functions. As one reviewer pointed out, however, our method does assume that the distributions would not otherwise materially change if people made smaller or larger adjustments. We used data from Experiment 1B to investigate this. First, p-values provided by Kolmogorov-Smirnov tests comparing the distribution of errors between payoff functions for each person ranged from .14 to 1.00. Thus aside from their obvious differences in location, these error distributions were statistically equivalent (at an alpha of .10) across payoff functions. Second, we assessed whether any differences in these error distributions, as indexed by the Kolmogorov-Smirnov test statistic D, were related to differences in participants' observed adjustments for each payoff function. This correlation was 0.15 ($p = .146$). In sum, differences between the error distributions across asymmetric payoff functions were minor and weakly related to differences in the observed adjustments.

[3] These three participants also reported significantly less motivation, $F(1, 182) = 3.64$, $p = .058$, and spent significantly less time on the task, $F(1, 183) = 8.16$, $p = .005$, than the other participants.

Table 1

Number of Lottery Tickets Earned, as a Function of Accuracy

		Estimate (relative to actual temperature)			
Block	Payoff	*More than 6°C below*	*6 to 0°C below*	*0 to 6°C above*	*More than 6°C above*
1	Equal reward for positive and negative errors	0	5	5	0
2	No reward for negative errors; reward for positive errors	0	0	5	2
3	No reward for positive errors; reward for negative errors	2	5	0	0

Note. Perfect estimates always earned five tickets. Because temperatures are less variable in Berkeley than in Pittsburgh, we reduced the cutoff from 6°C to 4°C in Experiments 1B and 2.

Table 2

Effects of Feedback on Beliefs and Overprecision in Experiment 2

Condition	*n*	Beliefs		Adjustments		Over-precision
		SD of SPD	Self-reported confidence	Observed	Normative	
No feedback	64	1.61 (0.56)	2.16 (0.93)	2.79 (1.88)	5.02 (1.47)	0.41 (0.51)
Authentic feedback	58	1.70 (0.43)	2.05 (0.91)	2.74 (1.07)	5.14 (1.52)	0.42 (0.28)
Exaggerated feedback	59	2.08 (0.46)	1.24 (0.90)	5.00 (2.66)	5.71 (1.70)	0.05 (0.51)
Overall	181	1.79 (0.53)	1.82 (1.00)	3.50 (2.23)	5.28 (1.59)	0.29 (0.48)

Note. Means reported with standard deviations in parentheses. SPD = Subjective probability

distribution.

Figure 1. A stylized representation of overprecision in judgment. The solid line plots a

hypothetical person's observed distribution of errors in response to payoffs which reward

overestimation; on average this person added +3 to her estimates. The double line identifies the

average adjustment which maximizes her expected payoffs, which in this case is +5. The percent

difference between the observed and normative adjustments represents overprecision. The

dashed line plots the person's belief about her errors implied by the observed adjustment. (Note,

we do not assume normality when calculating overprecision.)