

A Behavioral Demonstration of Overconfidence in Judgment

Albert E. Mannes¹ and Don A. Moore²

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

Psychological Science
24(7) 1190–1197
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797612470700
pss.sagepub.com



Abstract

Overprecision—an excessive confidence that one knows the truth—is both the most durable and the least understood form of overconfidence. This article outlines an approach to the study of overprecision that avoids some of the methodological problems of other approaches and better reflects the way uncertainty affects choices in everyday life. We measured the precision in judgment implied by people’s tendency to adjust their point estimates of an uncertain quantity in response to the costs of overestimating or underestimating the correct answer. The results revealed robust overprecision. People adjusted their estimates less than they should have given their actual knowledge, and this effect was driven by their subjective confidence.

Keywords

judgment, decision making, prediction, simulation

Received 6/4/12; Revision accepted 11/6/12

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 forms of overconfidence, the most robust and least understood is what Moore and Healy (2008) call overprecision—excessive confidence that one knows the truth. Overprecision has profound consequences, giving investors excessive confidence in the value of their investments (Daniel, Hirshleifer, & Subrahmanyam, 2001), leading physicians to gravitate too quickly to a particular diagnosis (Christensen-Szalanski & Bushyhead, 1981), and making people intolerant of dissenting views (Harvey, 1997; Harvey & Fischer, 1997).

Overprecision has proven remarkably resistant to debiasing by researchers using standard approaches to its study (Harvey, 1997; Soll & Klayman, 2004).¹ 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 (CIs) around a numerical estimate (e.g., the length of the Nile River). These 90% CIs include the truth as rarely as 30% of the time, which suggests that people have excessively precise views of what they know (McKenzie, Liersch, & Yaniv, 2008; Soll & Klayman, 2004).

One criticism of this approach is that it requires familiarity with probability and CIs—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 does not reflect the way overprecision affects people’s judgments in daily life. It is rare for people to have to specify CIs around a belief. When deciding what time to depart for a lunch meeting, for example, people do not consider a 90% CI around the travel time. Instead, if punctuality is important, uncertainty should lead them to depart earlier. Or because falling off a cliff is worse than missing the great view, people keep back from the precipice. Every day, in hundreds of different instances, people steer clear of metaphorical cliffs as a function of the cost of erring on each side and of uncertainty. In the research reported here, we manipulated these two factors to observe their effects on people’s judgments. The first benefit of our approach is that it does not require asking participants about probabilities, degrees of confidence, or CIs. The

Corresponding Author:

Albert E. Mannes, The Wharton School, University of Pennsylvania,
3730 Walnut St., Philadelphia, PA 19104
E-mail: albert.mannes@gmail.com

second benefit is that our approach parallels the way in which uncertainty affects everyday judgments.

Overview

Our approach builds on that of Mamassian (2008) by assessing 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 tickets in a lottery for a prize. The days were divided into three blocks, and each block featured a different payoff function. One block featured symmetric payoffs, in which participants earned lottery tickets if their estimates were correct or close to (either above or below) 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 of the truth or the other in daily life.

The participants in Experiment 1a, for instance, earned five lottery tickets in the block with symmetric payoffs if their estimates were within 6 °C of the correct answer and zero tickets otherwise (see Table 1). In the block with payoffs for overestimation, participants earned five tickets if they correctly guessed the answer or overestimated it by no more than 6 °C, earned two tickets if they overestimated it by more than 6 °C, and zero tickets if they underestimated it. The converse was true in the block with payoffs for underestimation: Participants earned five tickets if they correctly guessed the answer or underestimated it by no more than 6 °C, earned two tickets if they underestimated it by more than 6 °C, and zero tickets if they overestimated it.

For example, imagine a participant who is asked to guess the high temperature in Pittsburgh on March 10, 2006. The participant believes the most likely answer is 10 °C but also that any temperature in the range of 7 °C to 12 °C is possible. In the block with symmetric payoffs, the costs of overestimation and of underestimation are identical, so the participant should guess the most likely answer (10 °C) to try to earn five lottery tickets. In the blocks with asymmetric payoffs, the participant’s guess should account for his or her uncertainty about the correct answer and for the relative costs of overestimating and underestimating it. For instance, because the participant believes that the temperature could have been 11 °C or 12 °C that day, a guess of 10 °C may underestimate the correct answer. To earn five tickets in the block with payoffs for overestimation, the participant should adjust his or her estimate upward (e.g., +3 °C) to reduce the probability of underestimating the correct value. As indicated in Table 1, however, the participant will earn only two tickets instead of five if he or she overestimates too much. Thus, to earn the maximum of five tickets, the participant should guess something close to the correct answer but not less than it.

The confidence that participants have in their answers should affect the degree to which they adjust their estimates in the asymmetric-payoffs blocks. If, for example, the aforementioned participant is less confident and believes that temperatures between 5 °C and 15 °C were possible that day, he or she should make even larger upward adjustments (e.g., +5 °C) to reduce the probability of underestimation.

We made two predictions about people’s behavior in this task. First, we expected people to be overprecise: That is, we expected people to adjust their point estimates in the correct direction in response to the asymmetric payoffs but to do so insufficiently. Experiments 1a and 1b demonstrate this basic result. Second, we expected

Table 1. Payoff Structures in Experiment 1a

Block	Estimate (relative to actual temperature)			
	More than 6 °C below	Up to 6 °C below	Up to 6 °C above	More than 6 °C above
Symmetric payoffs: equal reward for underestimation and overestimation	0	5	5	0
Asymmetric payoffs: no reward for underestimation, reward for overestimation	0	0	5	2
Asymmetric payoffs: no reward for overestimation, reward for underestimation	2	5	0	0

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

people's confidence to moderate their overprecision. Specifically, we predicted that the more confidence people had in their point estimates, the less they would adjust them. We manipulated confidence in Experiment 2. (Data and stimuli for all three experiments are available at Don Moore's Web site: <http://learnmoore.org/mooredata/EJM/>.)

Experiment 1a

Method

We recruited 36 women and 28 men from Pittsburgh, Pennsylvania (mean age = 25.6 years), to participate in a study of judgment conducted at a local university. Participants were offered \$5 and tickets in a lottery for an iPod Shuffle in return for completing the study.

Participants first provided 90% CIs for the high temperature in Pittsburgh (in degrees Celsius or degrees Fahrenheit, at their choosing) on 24 randomly selected days in 2006 and 2007. These data enabled us to compare traditional measures of overprecision with our measure. Next, participants made point estimates of the high temperature on 72 days, which were divided into three blocks of 24 days, and received trial-by-trial feedback with the correct answers. Participants earned tickets according to the payoffs in each block, up to a maximum of 120 per block. Table 1 shows the payoffs for each block. The block with symmetric payoffs came first, followed by the two blocks with asymmetric payoffs; the order of the asymmetric-payoffs blocks was counterbalanced. Before starting each block, participants had to correctly answer three questions about the payoffs to demonstrate their attention and understanding. Moreover, the payoff function for the current block was summarized at the top of the computer screen throughout the trials. After completing their estimates, participants answered questions about themselves and their knowledge of Pittsburgh temperatures. After learning how many tickets they had earned, they were dismissed.

To measure overprecision in the point estimates, we calculated the ratio between the *observed adjustment* of each person's estimates in the asymmetric-payoffs blocks and a *normative adjustment* that would have 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. We expected a positive bias when participants were rewarded for overestimating the correct answer (i.e., a positive average adjustment in the block with payoffs for overestimation) and a negative bias when participants were rewarded for underestimating the correct answer (i.e., a negative average adjustment in the block with payoffs for underestimation).

To determine the normative adjustment for each person, we first calculated how many lottery tickets the person would have earned in each of the asymmetric-payoffs blocks for a range of additional adjustments to his or her observed estimates. Specifically, for each asymmetric-payoff block, we simulated the number of tickets that would have been earned for 201 additional adjustments to the observed estimates (viz., -50 °C, -49.5 °C, . . . $+49.5$ °C, $+50$ °C; note that an adjustment of $+0$ °C reproduces the observed estimates). We then averaged those adjustments for which the simulated earnings exceeded the participant's actual earnings. Finally, we added this number to the participant's observed adjustment to produce his or her normative adjustment for that block (cf. Lawrence & O'Connor, 2005).²

Overprecision was indexed by subtracting from 1 the ratio of each person's observed adjustment to his or her normative adjustment. If the observed adjustment was less than the normative adjustment, overprecision was greater than zero (the person was overprecise); 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 underprecise nor overprecise). For example, assume that a person's observed adjustment in the block with payoffs for overestimation was $+3$ °C and the simulations demonstrated that the person's earnings would have been higher with an average additional adjustment of $+2$ °C. This person's normative adjustment would be $+5$ °C, and his or her degree of overprecision would be 0.40 ($1 - 3/5$). This example is illustrated in Figure 1.

Results and discussion

Preliminary analysis. Two point estimates greater than 100 °C were coded as entry errors and excluded prior to the analysis. To check the manipulation of payoffs, we calculated the observed adjustment (in degrees Celsius) by each person for each block. The average adjustment in the block with symmetric payoffs was 0.69 , 95% CI = $[0.25, 1.13]$; in the block with payoffs for overestimation, the average adjustment was 4.93 , 95% CI = $[4.35, 5.50]$; and in the block with payoffs for underestimation, the average adjustment was -6.29 , 95% CI = $[-6.88, -5.70]$. Our manipulation of payoffs led people to adjust (i.e., bias) their estimates as expected.

Point estimates. Because differences between the two asymmetric-payoffs blocks were not of substantive interest, we averaged the observed and normative adjustments (reverse-coded as necessary) across these blocks for each person before calculating overprecision. The average level of overprecision ($M = 0.33$, $SD = 0.26$) was

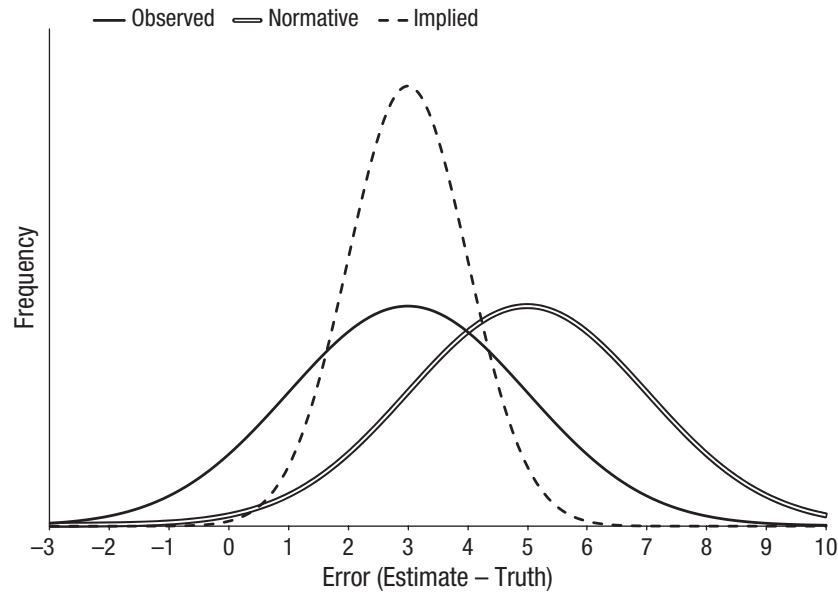


Fig. 1. A representation of overprecision in judgment. The solid line plots a hypothetical person's observed distribution of errors in response to payoffs that reward overestimation; on average, this person added +3 to her estimates. The double line identifies the average adjustment that would have maximized her expected payoffs, which in this case is +5. The percentage 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 that although we portray this person's errors as normally distributed, we did not make this assumption when calculating overprecision in the experiments.)

significantly different from 0, $t(63) = 10.16$, $p < .001$, Cohen's $d = 1.27$. Participants' actual earnings in the asymmetric-payoffs blocks ($M = 118.69$ tickets, $SD = 12.06$) were significantly less than what they would have earned absent their overprecision ($M = 125.05$ tickets, $SD = 8.99$), $t(63) = 19.49$, $p < .001$, $d = 0.60$.

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

Confidence intervals. On average, participants' 90% CIs contained the correct answer 49% of the time ($SD = 24\%$). Thus, this study replicated the overprecision in CIs found in prior research. There was also correspondence in the measures of overprecision using our method and using CIs. Fifty-four of the participants were overprecise in both their CIs and their point estimates, 7 were overprecise only in their CIs, and 3 were overprecise only in their point estimates. In addition, individuals making the smallest adjustments to their point estimates constructed the narrowest 90% CIs ($r_s = .43$, $p < .001$).

Experiment 1b

Presenting the symmetric-payoffs block first in Experiment 1a provided participants with extensive feedback about their actual knowledge, which enabled them to make informed judgments about how to adjust their estimates in the asymmetric-payoffs blocks. However, it is possible that participants found it difficult to switch from making their best estimate of the temperature in the context of symmetric payoffs to making biased estimates when payoffs were asymmetric. If this was the case, the results of Experiment 1a would overstate the extent of overprecision. With this concern in mind, in Experiment 1b we counterbalanced the order of the symmetric- and asymmetric-payoff blocks.

Method

We recruited 60 women and 42 men from Berkeley, California (mean age = 20.7 years), to participate. Each was offered \$10 and a chance to win a \$50 gift card from an online retailer. Participants guessed the high temperature in Berkeley (in degrees Celsius or degrees Fahrenheit, at their choosing) on randomly chosen days from 2006 and 2007. The procedure and measures were identical to those of the prior experiment, with three exceptions. First, we

counterbalanced the order of the symmetric- and asymmetric-payoff blocks. Second, participants provided their 90% CIs after finishing all three blocks of point estimates. Third, the cutoffs in the payoff functions were changed from 6 °C to 4 °C.

Results and discussion

The average adjustment (in degrees Celsius) was -1.09 , 95% CI = $[-1.34, -0.84]$, in the block with symmetric payoffs; 2.79 , 95% CI = $[2.46, 3.12]$, in the block with payoffs for overestimation; and -2.84 , 95% CI = $[-3.12, -2.57]$, in the block with payoffs for underestimation.

The average level of overprecision ($M = 0.40$, $SD = 0.26$) was significantly different from 0, $t(101) = 15.82$, $p < .001$, $d = 1.57$. Overprecision was marginally higher for the 53 participants who started with asymmetric payoffs ($M = 0.45$, $SD = 0.26$) than for the 49 participants who started with 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.

On average, participants' 90% CIs contained the correct answer 70% of the time ($SD = 18\%$). Seventy-five of the participants were overprecise in both their CIs and their point estimates, 8 were overprecise only in their CIs, 18 were overprecise only in their point estimates, and 1 was overprecise in neither task. As in the prior study, individuals making the smallest adjustments to their point estimates constructed the narrowest 90% CIs ($r_s = .38$, $p < .001$).

Experiment 2

Experiments 1a and 1b produced strong evidence of overprecision in judgments. In Experiment 2, we investigated the role of confidence as a moderator of people's adjustments of point estimates. We manipulated confidence directly by varying the feedback participants received about the accuracy of their estimates (cf. Arkes, Christensen, Lai, & Blumer, 1987; Stone & Opel, 2000). Specifically, participants received no trial-by-trial feedback about their accuracy, authentic feedback (as in Experiments 1a and 1b), or feedback that exaggerated their errors. We expected that the overprecision found in Experiments 1a and 1b would be replicated in the no-feedback and authentic-feedback conditions and that participants who received exaggerated feedback would show the least overprecision.

Method

We recruited 107 women and 77 men from Berkeley, California (mean age = 20.5 years), to participate. Each

was offered \$10 and a chance to win a \$25 gift card from an online retailer. The procedure was the same as that of Experiment 1a, with three exceptions. First, the cutoffs in the payoff functions were changed from 6 °C to 4 °C. Second, after completing a block of 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. In the *exaggerated-feedback* condition, participants received manufactured trial-level feedback about their errors. The feedback was exaggerated by telling participants that their errors were approximately 2.5 times greater than they actually were (e.g., if the actual error was +2 °C, the participant was told that the error was +5 °C).

Third, after making their point estimates in the symmetric-payoffs block, all participants provided subjective probability distributions of their errors using the subjective-probability interval-estimate technique (Haran, Moore, & Morewedge, 2010). They indicated how many of their errors would fall into each of nine intervals for every 100 temperature estimates. (In degrees Celsius, these nine intervals were less than -11 , -10 to -8 , -7 to -5 , -4 to -2 , -1 to $+1$, $+2$ to $+4$, $+5$ to $+7$, $+8$ to $+10$, and greater than $+11$.) After assigning a frequency to each interval, participants completed the asymmetric-payoffs blocks. Participants then answered questions about themselves, their understanding of the task, their motivation, and their confidence. Finally, they learned how many tickets they had earned and were dismissed.

Overprecision was operationalized in the same way as in the prior experiments. We collected two measures of participants' confidence: one implicit and one explicit. The implicit measure was the subjective probability distribution of their errors that participants provided after completing the symmetric-payoffs block. 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 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 greater than 100 °C in absolute value were dropped as data entry errors. In addition, 3 participants who made constant and extreme estimates in the asymmetric-payoffs blocks (e.g.,

–18 °C for Berkeley, CA) were excluded from the analysis, which left a final sample of 181 participants.³

The average adjustment (in degrees Celsius) was –0.29, 95% CI = [–0.47, –0.10], in the block with symmetric payoffs; 3.78, 95% CI = [3.40, 4.15], in the block with payoffs for overestimation; and –3.21, 95% CI = [–3.63, –2.80], in the block with 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). Our manipulation of confidence was successful: A multivariate omnibus test rejected the possibility that feedback had no effect on confidence, $F(4, 178) = 13.20, p < .001$. As Table 2 indicates, participants in the no-feedback condition were most confident, followed by those in the authentic-feedback condition, followed by those in the exaggerated-feedback condition. This was true for both the subjective probability distributions, $F(2, 178) = 16.33, p < .001, \eta^2 = .15$, and self-reported confidence, $F(2, 178) = 18.28, p < .001, \eta^2 = .17$.

Point estimates. The average level of overprecision ($M = 0.29, SD = 0.48$) was significantly different from 0, $t(180) = 8.31, p < .001, d = 0.62$. Participants' total earnings in the asymmetric-payoffs blocks ($M = 116.40$ tickets, $SD = 18.77$) 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 who received exaggerated feedback were significantly less overprecise than those who received 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 0, $t(58) = 0.72, p = .47, d = 0.09$.

Participants displayed overprecision in this task by insufficiently adjusting their point estimates relative to their normative benchmarks (see Table 2 for participants' average observed and normative adjustments for each

condition). We have argued that such adjustments are influenced by confidence: People who are less confident in their expertise adjust their estimates more than people who are more confident do. The data supported this claim. Specifically, participants with wider subjective probability distributions made larger adjustments to their subsequent point estimates ($r = .37, p < .001$) and were less overprecise ($r = -.28, p < .001$) than participants with narrower subjective probability distributions.

Confidence intervals. On average, participants' 90% CIs contained the correct answer 63% of the time ($SD = 22\%$). This overprecision did not differ by condition, $F(2, 178) = 0.91, p = .405$. Of the 181 participants, 121 were overprecise in both their CIs and their point estimates, 37 were overprecise only in their CIs, 17 were overprecise only in their point estimates, and 6 were overprecise in neither task. Individuals who made the smallest adjustments to their point estimates constructed the narrowest 90% CIs ($r_s = .18, p = .018$).

Discussion

The three experiments we have reported provide evidence of overprecision in judgment in a paradigm that does not require participants to understand probabilities or CIs and that reflects 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 that assume too much about respondents' statistical sophistication.

Yet two concerns about this method persist. First, participants may have simply been insensitive to the asymmetric payoff structure. Anticipating this possibility, we took procedural measures to ensure that participants understood and were aware of the payoffs. Moreover, participants adjusted 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

Table 2. Mean Confidence, Adjustments, and Overprecision in Experiment 2

Condition	<i>n</i>	Confidence		Adjustment		Overprecision
		Implicit	Self-reported	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: Standard deviations are given in parentheses. The implicit measure of confidence was the standard deviation of the participant's subjective probability distribution, so lower values indicate higher confidence.

to asymmetric payoffs, see Goodwin, 2005; Lawrence & O'Connor, 2005; Weber, 1994).

Second, even if participants understood the payoff structures, they may have insufficiently adjusted their estimates because they were anchored by their best estimates of the answers (Block & Harper, 1991). In fact, verbal protocols from a supplementary study confirmed that participants often started with a best guess and adjusted from there. However, although anchoring processes may have been at work, they cannot explain why (a) participants with less expertise made greater adjustments in Experiment 1a, (b) participants with less confidence made greater adjustments in Experiment 2, and (c) overprecision was eliminated in Experiment 2 when people received unfavorable feedback about their accuracy. In sum, when asymmetric incentives lead people to adjust their responses to one side or the other of a best guess, the amount of adjustment depends critically on how confident they 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 presented here 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 CI estimation (Russo & Schoemaker, 1992). Nevertheless, it offers researchers an alternative to self-reports that rely on having to explain probabilities or CIs.

The consequences of overprecision are profound. People frequently cut things too close—arriving late, missing planes, bouncing checks, or falling off one of the many “cliffs” that present themselves in daily life. 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 evidence we have presented here will prove useful in elucidating the underlying sources of people's excessive faith in the accuracy of their own judgment.

Author Contributions

A. E. Mannes and D. A. Moore developed the theory and design used in all three experiments. A. E. Mannes programmed the experiments and conducted the analysis of the data. D. A. Moore collected the data. A. E. Mannes and D. A. Moore contributed equally to the drafting and revision of the manuscript. Both authors approved the final version of the manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. There has been some success at debiasing overprecision. Elicitation methods that call explicit attention to a large range of response values can reduce overprecision (see Haran, Moore, & Morewedge, 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 these distributions would not otherwise materially change if people made smaller or larger average adjustments. We used data from Experiment 1b to investigate the validity of this assumption. First, p values provided by Kolmogorov-Smirnov tests comparing the distribution of errors between the asymmetric-payoff blocks for each person ranged from .14 to 1.00. Thus, aside from their obvious differences in location, these error distributions were statistically equivalent (at $\alpha = .10$) between payoff conditions. 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 condition. This correlation was .15 ($p = .146$). In sum, differences in the error distributions between the two asymmetric-payoff blocks were minor and weakly related to differences in the observed adjustments.

3. These 3 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.

References

- Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*, 133–144. doi:10.1016/0749-5978(87)90049-5
- Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes*, *49*, 188–207. doi:10.1016/0749-5978(91)90048-x
- 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. *Journal of Forecasting*, *14*, 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*, 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*, 1–73.
- Daniel, K. D., Hirshleifer, D., & Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *Journal of Finance*, *56*, 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*, 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, The Netherlands: 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, England: Wiley.
- Goodwin, P. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research*, *163*, 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*, 199–207.
- Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*, 467–476.
- Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Sciences*, *1*, 78–82. doi:10.1016/s1364-6613(97)01014-0
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.
- Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting*, *21*, 3–14. doi:10.1016/j.ijforecast.2004.02.003
- Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science*, *19*, 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*, 179–191. doi:10.1016/j.obhdp.2008.02.007
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*, 502–517. doi:10.1037/0033-295X.115.2.502
- Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review*, *33*, 7–17.
- Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 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*, 282–309. doi:10.1006/obhd.2000.2910
- Tversky, A., & Kahneman, D. (1993). Probabilistic reasoning. In A. I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 43–68). Cambridge, MA: 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. *Psychological Bulletin*, *115*, 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1167–1175. doi:10.1037/0278-7393.30.6.1167