Judgment and Decision Making (2025), **vol. 20**:e25 1–30
doi:10.1017/jdm.2025.11

EMPIRICAL ARTICLE

Is overconfidence an individual difference?

Sophia Li[1], Randall Hale[2], and Don A. Moore[1]

[1]Haas School of Business, University of California Berkeley, Berkeley, CA, USA and [2]Department of Psychology, UCLA, Los Angeles, CA, USA

Corresponding author: Sophia Li; Email: sophia-li@berkeley.edu

Received: 6 January 2025; **Revised:** 26 March 2025; **Accepted:** 28 March 2025

Keywords: overconfidence; individual differences; overestimation; overplacement; overprecision

Abstract

Some scholars have treated overconfidence as an individual difference—that is, assuming the tendency to be overconfident is stable within a person and differs meaningfully from person to person. We question this assumption. We investigate consistency within individuals between its three forms—overestimation, overplacement, and overprecision—in multiple domains (Study 1a and 1b), at multiple times (Study 1b and 2), and with multiple measures (Study 3a and 3b). We find mixed evidence of trait-like consistency. We do find some evidence of within-individual stability across domains and time points. However, we find little consistency across different measures of the same form of overconfidence—specifically overprecision. Instead, we find more consistent evidence that overconfidence varies situationally and contextually.

1. Introduction

Overconfidence is one of the most prominent and pervasive of all cognitive biases (Bazerman and Moore, 2012; Kahneman, 2011). It is the first bias that Daniel Kahneman said he would eliminate if he could (Shariatmadari, 2015). Considering the ubiquity and impact of overconfidence, it is natural to ask whether certain people are more overconfident than others. Lawson et al. (2023) report 'strong evidence for stable, individual differences in overconfidence'. Binnendyk and Pennycook (2024) concluded that 'some individuals consistently overestimate their abilities' and that this overestimation predicted their endorsement of conspiracy theories, overclaiming, and receptivity to bullshit.

These claims are bolstered by studies reporting relationships between overconfidence and other traits. For example, researchers have claimed that 'narcissists have higher levels of confidence' (O'Reilly and Hall, 2021); that 'men are more overconfident than women' (Bengtsson et al., 2005); and that some CEOs are more overconfident than others (Malmendier and Tate, 2005). If some people are consistently more overconfident, it follows that certain professions, such as entrepreneurship, might select the most overconfident, whereas other professions, such as reinsurance or disaster preparedness, might select against them (Hogarth and Karelaia, 2012; Larkin and Leider, 2012). All of this is premised on the claim that overconfidence is a stable trait that varies between people—that some people are consistently more overconfident than others.

We ask whether overconfidence is consistent within a person. Put another way, is overconfidence more similar to intelligence or risk preferences? General intelligence is generally accepted as an individual difference; one's intelligence is somewhat consistent from one situation to the next

(Spearman, 1961), though scholars have identified multiple types of intelligence (Gardner and Hatch, 1989) and some continue to argue that it is not a trait but rather an interaction between a person and situation (Sternberg, 2021).

Conversely, evidence suggests that risk preferences are not stable across situations (Weber et al., 2004). Some skydivers invest conservatively and some gamblers drive cautiously (Weber and Johnson, 2009). Nevertheless, some researchers have treated risk preferences as if they represented traits (Holt and Laury, 2002). We ask whether overconfidence is more like intelligence or risk preferences; in other words, whether there is evidence of some general overconfidence trait that is observable across situations and tasks.

1.1. Prior findings

Overconfidence is being more confident than is justified or deserved (Moore and Dev, 2017). We distinguish overconfidence from constructs such as confidence and optimism; overconfidence compares subjective beliefs with the truth. For example, someone who is *confident* in their performance on a math test may not be *over*confident if, in reality, they perform very well; if they perform badly despite their confidence, we would call them overconfident. Scholars have examined three primary forms of overconfidence, each form distinguished by its truthful benchmark (Moore and Schatz, 2017): overestimation is thinking that you are better than you are (e.g., thinking that you answered 5 questions correctly when you only got 3); overplacement is the exaggerated belief that you are better than others (e.g., thinking that you were in the top 10% of the class when you were in the bottom 10%); overprecision is being too sure you know the truth (e.g., being 100% certain that you answered 5 questions correctly when you answered 8). Although one could conceive of overprecision as a specific instance of overestimation—overestimating your ability to identify the truth—we distinguish them by operationalizing overprecision as relating to certainty in judgment, and overestimation as relating to performance on a task.

What would it mean for there to be some general overconfidence trait? Overconfidence would have to persist across forms, times, or measures within individuals. Empirically, we would expect to observe correlations between overconfidence measures across tasks and time, at least within the same form of overconfidence. We would also expect that at least one related stable trait measure would predict overconfidence in at least one form and potentially point toward why some people might be more overconfident than others. For example, if overconfidence was at least partially caused by a failure to evaluate contradictory evidence, we might expect actively open-minded thinking to correlate with overprecision when contradictory evidence is within reach. Actively open-minded thinkers are better-calibrated forecasters (Mellers et al., 2015), while those who are less open-minded tend to favor information that confirms preexisting beliefs (Stanovich et al., 2013), that is, are susceptible to myside bias. It could also be the case that meta-reasoning ability—the ability to monitor one's own reasoning, including feelings of being right or wrong (Ackerman and Thompson, 2017)—is related to general intelligence or cognitive ability, in which case we might expect more intelligent people to also be better calibrated. Previous research has suggested that people with higher reasoning abilities are less susceptible to common heuristics and biases (Jackson et al., 2016) so we might expect them to also be less susceptible to overconfidence. If, on the other hand, individuals construct confidence judgments at the moment, then situational influences will prevail. For example, people will think they are better than others on easy tasks but worse than others on hard tasks, inserting situational variation that diminishes any stability from an individual difference (Moore and Small, 2007).

We take two approaches to reviewing the published literature. First, we ask whether overconfidence correlates with individual traits or with demographics; such evidence would point towards overconfidence resulting from within-individual stability. Second, we review studies that consider correlations between similar measures of overconfidence across different contexts. If overconfidence is an individual difference, we would expect to see stability within individuals across different contexts, such as domains and time. Our results leave us concerned about the strength and consistency of the evidence.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

We start by considering the relationship between overconfidence and demographics. Perhaps the most high-profile claim is that men are more overconfident than women; indeed, stereotypes hold that men are excessively certain of their views (which is why they 'mansplain'). Barber and Odean (2001) find that men invest more than women but have lower returns, and Niederle and Vesterlund (2007) find that men overplace their performances on arithmetic tasks more than women. However, it seems that these gender differences may be task-dependent (Beyer, 1990; Dahlbom et al., 2011; Lundeberg et al., 1994); for example, Exley and Kessler (2022) find gender gaps in self-evaluations on male-typed subjects (math and science) but not in female-typed subjects (verbal skills). Further, several studies have failed to replicate gender differences in overconfidence even on similar finance-related tasks (e.g., Acker and Duck, 2008; Deaves et al., 2010). Finally, while there is some evidence that certainty might increase with age (Crawford and Stankov, 1996), this does not seem to hold for overestimation or overplacement (Prims and Moore, 2017), and may be dependent on specific measures of precision (Hansson et al., 2008).

Similarly, evidence is inconsistent on whether personality traits correlate with overconfidence. For example, some scholars have found a positive association between narcissism and overconfidence (Ames and Kammrath, 2004; Binnendyk and Pennycook, 2024; Campbell et al., 2004), perhaps because narcissists possess a grandiose self-assessment which makes it difficult for them to admit their own shortcomings or uncertainty; however, others have failed to replicate this claim (Moore and Swift, 2010). Results testing whether thinking styles are related to overconfidence are also contradictory; research on conspiracy beliefs (Pennycook et al., 2022) and bullshit receptivity (Littrell et al., 2021) report that these traits correlate with overestimation of cognitive abilities, while Hoppe and Kusterer (2011) report no predictive effect of cognitive reflection on overestimation. Recent research on intellectual humility reports inconsistent relationships with overconfidence that are measure-dependent for both intellectual humility and overconfidence (Bowes et al., 2024). Spiller (2024) notes that some of these inconsistencies may be due to correlations between measures of overconfidence and ability. The results are mixed, to say the least, and the literature has not achieved consensus around any one of these traits consistently predicting overconfidence. Publication bias would also predict that null relationships between overconfidence and trait measures are less likely to be present in the published literature than positive ones.

The second kind of evidence we review asks whether similar measures of overconfidence correlate with each other across performance domains. This is a low bar, which only seeks test-retest reliability. While there seems to be evidence for within-individual stability of confidence (Pallier et al., 2002; Stankov and Crawford, 1996), evidence for overconfidence is less consistent. The strongest case for overestimation comes from Klayman et al. (1999)'s assessment of overestimation of performance on various tests, such as people's estimates of life expectancies in various countries around the world; they report within-person correlations in the neighborhood of .5. However, West and Stanovich (1997) report lower correlations between .07 and .24 between overestimation on a general knowledge test and overestimation on predicted performance on a motor task in which participants slid pennies into a target zone. Bornstein and Zickafoose (1999) found a correlation of 0.3 between overestimation on a general knowledge task and overestimation on an eyewitness memory test. The strongest evidence of test-retest reliability for overprecision is Moore and Healy's (2008) high reliability ($\alpha = 0.95$) within participants on a particular measure of the certainty (overprecision) with which they estimated quiz scores; however, they found lower reliability for overestimation ($\alpha = 0.21$) and overplacement ($\alpha = 0.29$).

Given the inconsistencies in prior research, the present research confronts at least three issues. First, we test domain generality. If there are robust individual differences in overconfidence, then those differences must be generalized across domains. Second, we test the persistence of overconfidence over time. Most of the studies examining domain generality measured overconfidence at a single time point. There are a few exceptions; Glaser et al. (2005) find weak pairwise correlations between overprecision on stock market forecasting tasks spaced 2 weeks apart; Jonsson and Allwood (2003) report consistent confidence calibration on verbal and visual reasoning tests spaced 2 weeks apart. In addition, data from Massey et al. (2011) reveal modest correlations from week to week in sports fans' optimism about their

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

team's chances of winning, though these relationships were not the focus of the article. Aside from these studies, there is little data on whether overconfidence persists over time.

The third issue arises from a basic omission in the published literature: Do different measures of the same form of overconfidence even correlate with one another? The best hope to find a stable individual difference would seem to be overprecision, given the alpha reliability of .95 in Moore and Healy's (2008) data. However, the question remains whether this reliability generalizes to other measures of overprecision. Researchers have employed many measures of overprecision including confidence interval widths (Langnickel and Zeisberger, 2016), Likert-scale confidence (Brewer and Sampaio, 2012), and belief distributions (Haran et al., 2010). If there are reliable differences between people in their overconfidence, different approaches to measuring the same construct must correlate with one another if each measure is reliable and valid. To our knowledge, research has not tested correlations between various measures of precision. If there are several valid approaches to eliciting certainty, then assessing the degree to which different measures correspond with one another is crucial to understanding the underlying phenomenon.

1.2. Overview of the present research

Study 1a assesses the within-individual stability of each form of overconfidence across three different tasks. Study 1b is a replication of Study 1a with the same participants as in Study 1a, 10 months later; thus, Study 1b serves the dual purpose of measuring test-retest reliability and the persistence of overconfidence across time. Study 2 assesses the within-individual stability of each form of overconfidence on a similar task at two different time points with less attrition than Studies 1a and 1b. Studies 3a and 3b test the degree to which different measures of overprecision correlate with one another. In sum, these studies test the stability of overconfidence the within-individual between different task domains, time points, and measures.

In addition to measures of overconfidence, we also measure the following traits (see Table 1 for a list of measures by study for a summary of our measured correlations with overconfidence): gender, age, the Big Five personality traits (Schaefer et al., 2004), narcissism (O'Reilly and Hall, 2021), intellectual humility (Krumrei-Mancuso and Rouse, 2016), actively open-minded thinking (Haran et al., 2013), need for cognition (Lins de Holanda Coelho et al., 2020), and need for cognitive closure (Webster and Kruglanski, 1994). We selected some of these because of existing published claims, e.g., on gender differences in overconfidence (Niederle and Vesterlund, 2007). We selected other traits due to their seeming conceptual overlap with overconfidence. Scholars have theorized that overprecision is a result of error neglect, that is, simply not knowing the ways in which one is wrong (Moore, 2023), and empirical evidence has suggested that forcing people to consider alternative possibilities—that is, ways in which they are wrong—can reduce overprecision (Koriat et al., 1980; Walters et al., 2017). If actively open-minded thinking is a 'summary measure. . .on the decisions to remain open to further thinking' then one might speculate that actively open-minded thinkers who are more likely to weigh contrary evidence would also be more aware of the possibilities of contrary evidence, and therefore, be less likely to overestimate their performance and be less overprecise. Conversely, people who have a high need for cognitive closure and value identifying one specific answer might be more overprecise.

Our article makes several contributions. First, we test the consistency of different forms of overconfidence across task domains. Second, we examine the intertemporal consistency of all three forms of overconfidence. Third, we elicit several different measures of overconfidence within the same study. Fourth, we collect a comprehensive (though non-exhaustive) suite of trait measures throughout our studies for which there are mixed or reasonably hypothesized claims of relationships with overconfidence.

Our data are correlational. We offer our own interpretation of the correlations, but we invite our readers to come to their own conclusions. We would offer a few numbers to put our results in context: Mischel's (1968) conclusion that correlations between behaviors in different situations below .4 should be interpreted as poor consistency; Fleeson and Gallagher's (2009) observations of the correlation

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 1. *Summary of confidence and trait measures collected by study.*

Study	Generality	Domain(s)	Confidence measures	Trait measure
1a	Domain	Sports (MLB) forecasting, Raven's Progressive Matrices, weight-guessing	Overestimation, overplacement, overprecision (subjective probability distribution)	Narcissism, actively open-minded thinking
1b	Domain Test-retest	Sports (MLB) forecasting, Raven's Progressive Matrices, weight-guessing, fuzzy image identification (GOT)	Overestimation, overplacement, overprecision (subjective probability distribution	Narcissism, actively open-minded thinking
2	Time	Sports (NFL) forecasting	Overestimation, overplacement, overprecision (subjective probability distribution), overestimation (alternate)	Narcissism, overconfidence test, actively open-minded thinking
3a	Measures of certainty	Fuzzy image identification (GOT)	Overestimation, overplacement, Likert, bet, 90% confidence interval, subjective probability distribution	Actively open-minded thinking, intellectual humility (independence of ego and intellect), intellectual humility (lack of intellectual overconfidence), overconfidence test, gender, age
3b	Measures of certainty	Sports (NBA) forecasting	Likert, bet, 90% confidence interval, subjective probability distribution, numeric probability	Big Five, actively open-minded thinking, intellectual humility (independence of ego and intellect), intellectual humility (lack of intellectual overconfidence), narcissism, need for cognition, need for cognitive closure

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

between personality traits and their behavioral manifestations around .4; the interpretation of Pearson correlation coefficient of .3 for individual differences using personality measures as 'weak' (Evans, 1996) or 'small' (Cohen, 1988) to 'moderate' (Funder and Ozer, 2019; Gignac and Szodorai, 2016).

1.3. Transparency and openness

We preregistered all of our studies and reported deviations from those preregistrations in the main text. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. All preregistered analyses, including secondary ones, are reported either in the main text or in the Supplementary Material. The Supplementary Materials, data, and code for all studies are available online at https://osf.io/tb2me/?view_only=77b881f0b92647689270ed3b485366f9.

2. Study 1a: Overconfidence across domains

Study 1a tests whether each of the three forms of overconfidence is consistent across domains. Specifically, we measured each of the three forms of overconfidence in participants' performance in three different domains: predicting winners of Major League Baseball games, Raven's Progressive Matrices, and a weight-guessing task. Participants completed a ten-item test in each of the three domains; then, they reported their confidence about their own performance (estimation and overplacement), others' performance (placement), and their certainty about the score distribution of all participants (precision). We would interpret high correlations between overconfidence measures across tasks and forms of overconfidence as evidence in favor of a general overconfidence trait. Further, we elicited measures of narcissism and actively open-minded thinking; we would interpret positive correlations between narcissism and measures of overconfidence—particularly overestimation and overplacement if narcissists have excessively positive views of themselves and their skills—and negative correlations between actively open-minded thinking and overconfidence (most likely overprecision), as evidence in favor of a general overconfidence trait.

2.1. Method

We preregistered this study at https://aspredicted.org/MZ5_5ZX on July 12, 2023 before data collection began on July 13, 2023.

2.1.1. Participants

We noted in the consent form and advertisement on Cloud Research that we were looking for Major League Baseball (MLB) fans to complete the survey. In addition, participants had to pass three screening questions at the beginning of the survey. First, we asked participants 'Do you identify as an MLB fan?' and stopped those who answered 'No' from completing the survey. In addition, we asked participants two multiple-choice questions about the champion and runners-up of the prior year's World Series; only those who answered both correctly could proceed. These screening criteria excluded 118 potential participants. We selected this sports forecasting context based on a pretest (see the Supplementary Material), where laypeople suggested the sports forecasting context as most likely to reveal stable overconfidence within individuals.

Four hundred and two participants from CloudResearch passed the screening criteria. We excluded nine participants who straightlined (answered every item with the same response) the actively open-minded thinking scale and 12 participants whose reported estimates for the average of other scores' differed by more than 2 from the mean of their reported subjective probability distributions for the same estimate, leaving us with a final sample of 381 participants. Our participants were 46.19% male, 53.28% female, 0.52% other gender; 70.6% White / 6.04% Asian / 10.76% Black / 6.3% Hispanic /

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

6.3% other race; M_{age} = 35.16, SD_{age} = 10.46; with a median education level of a Bachelor's degree in college. Participants earned $3.00 for the study, which took 19.34 min on average.

This study required two rounds of data collection because we failed to attain the planned sample by the time the forecasted games began. The first round of data came in between July 13 and July 21, 2023 with participants forecasting MLB games that took place on July 21. The second round came in between July 27 and July 31, 2023 with participants forecasting MLB games that took place on August 4.

2.1.2. Procedure

Participants each completed three tasks, presented in a randomized order: forecasting the outcomes of 10 regular-season MLB games that would take place on July 27 or August 4, 2023 by predicting the winner (binary choice) of each game; 10 of the eight-choice questions from the short form of the Advanced Progressive Matrices (specifically questions 3, 12, 15, 16, 18, 21, 22, 28, 30, and 34; Bors and Stokes, 1998), a more difficult version of Raven's Progressive Matrices (RPM); and a ten-question weight-guessing task (WGT) in which participants gave point estimates (in pounds) of the weights of photographed individuals.

After each task, participants reported confidence. Specifically, they estimated their score (out of 10), the percentage of other participants in the study they believed they outperformed (0–100%), the average score for all participants (out of 10), and the distribution of scores for all participants, measured via a subjective probability distribution (SPD). Participants then completed two trait measures: narcissism (NPI) and actively open-minded thinking (AOT). Finally, participants reported gender, age, education, and ethnicity.

2.1.3. Measures

Actual Scores (Accuracy). Participants' actual scores (M_{MLB} = 5.02, SD = 1.55; M_{RPM} = 4.38, SD = 2.15; M_{WGT} = 1.85, SD = 1.54) were calculated as follows: for MLB forecasting, participants received 1 point for every game whose winner they predicted correctly; for Raven's Progressive Matrices, participants received 1 point for every question answered correctly; for weight-guessing, participants received 1 point for every weight was within 10 pounds of their guess.

Confidence (Estimation). After each test, participants estimated their scores out of 10, for example, 'How many questions out of 10 do you think you got correct on this visual reasoning test?' (M_{MLB} = 5.49, SD = 1.72; M_{RPM} = 4.21, SD = 2.06; M_{WGT} = 5.40, SD = 2.16). For the weight-guessing task, we also informed participants that 'a guess counts as 'correct' if your estimate falls within 10 lbs (above or below) the person's actual weight'.

Confidence (Placement). For our primary measure of overplacement, we asked participants to estimate the average score of all participants in the study, e.g., 'Several hundred other participants also completed this test. What do you think the average score of all participants in the study will be?' (M_{MLB} = 5.36, SD = 1.35; M_{RPM} = 4.74, SD = 1.48; M_{WGT} = 5.28, SD = 1.46). We calculated placement, that is, the degree to which participants think they are better than others, by subtracting participants' estimates of the average of others scores from their estimates of their own scores (M_{MLB} = 0.13, SD = 1.87; M_{RPM} = −0.53, SD = 2.12; M_{WGT} = 0.13, SD = 1.83). Our secondary measure of placement was based on an estimate of percentile and directly asked participants, 'What percentage of other participants in this study do you think you scored higher than?' (M_{MLB} = 42.45, SD = 21.30; M_{RPM} = 38.46, SD = 25.18; M_{WGT} = 41.97, SD = 23.23).

Confidence (Precision). We asked participants to complete a subjective probability distribution for the distribution of all participants' scores on each test, e.g., 'For each row, estimate the percentage of other participants who scored that many points on this MLB Prediction test' (from 0 to 10). Participants then adjusted a series of 11 slider bars to indicate the percentage for each score. We calculated variance from participants' subjective probability distributions of the score distribution by (a) calculating the mean of the distribution by summing the product of each score and its probability, and (b) squaring the distance from the mean for each score and summing each squared distance with its associated probability (M_{MLB} = 5.58, SD = 2.67; M_{RPM} = 5.24, SD_{RPM} = 2.65; M_{WGT} = 5.40, SD = 2.65).

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Overestimation. For each of the three tests, we calculated overestimation as the participant's estimated scores minus their actual score (M_{MLB} = 0.47, SD = 2.22; M_{RPM} = −0.18, SD = 2.265 M_{WGT} = 3.56, SD = 2.69). Calculating overestimation by subtracting a participant's actual score follows prior literature, e.g., Moore and Healy (2008).

Overplacement. For each of the three tasks, we calculated overplacement as (estimated own score − estimate of mean score of all participants) − (actual own score − actual mean score of all participants) (M_{MLB} = 0.12, SD = 2.36; M_{RPM} = −0.58, SD = 2.39; M_{WGT} = 0.10, SD = 2.41). Calculating overplacement using this formula follows prior literature, e.g., Logg et al. (2018). We also calculated a secondary (percentile-based, sometimes known as direct overplacement) measure of overplacement as (estimated own percentile rank − actual own percentile rank) (M_{MLB} = 1.34, SD = 36.18; M_{RPM} = −5.78, SD = 33.16; M_{WGT} = 1.29, SD = 38.53). For the secondary measure of placement and overplacement, we include results but do not discuss them. In general, their inter-task and inter-time correlations are weaker than those of the indirect measure based on average score estimates.

Overprecision. For each of the three tasks, we calculated overprecision by calculating (actual variance of participants' scores − variance of subjective probability distribution) (M_{MLB} = −3.18, SD = 2.67; M_{RPM} = −0.61, SD = 2.65; M_{WGT} = −3.02, SD = 2.65)[1]. Calculating precision and overprecision based on the variance of a subjective probability distribution follows previous literature (Haran et al., 2010; Moore, Carter, et al., 2015). Although there are many possible measures of precision and overprecision, as employed in Studies 3a and 3b, we selected subjective probability distribution because of evidence suggesting these histogram elicitations result in better-calibrated judgments than point estimates or confidence intervals (Goldstein and Rothschild, 2014). Further, this measure allows us to capture overprecision (not just precision) with only one measure per task, as compared to methods such as several different estimates and confidence intervals to measure calibration.

Narcissism. Participants completed the 16-item Narcissistic Personality Inventory (Ames et al., 2006); each item presented two contrasting statements and participants were instructed to choose 'the statement (left or right) that best describes you,' for example, 'I really like to be the center of attention' versus 'It makes me uncomfortable to be the center of attention,' α = 0.73, M = 12.40, SD = 2.96.

Actively open-minded thinking. Participants completed the 7-item actively open-minded thinking scale (Haran et al., 2013), for example, 'People should take into consideration evidence that goes against their opinions', 1 = 'strongly disagree' to 5 = 'strongly agree' α = 0.74, M = 3.76, SD = 0.53.

2.2. Results

2.2.1. Correlations across domains

This study's key analyses are the pairwise correlations of each type of overconfidence between tasks (see Table 2). The average (across task domains) inter-task correlations for each form of overconfidence were relatively weak on overestimation (0.12) and overplacement (0.10), and higher on overprecision (0.68). Despite the correlations being relatively low in absolute value, several pairwise correlations were significantly >0. Overestimation on the MLB forecasting task was correlated positively with both overestimation on Raven's Progressive Matrices, r = .22, p < .001, and overestimation on the weight-guessing task, r = .15, p = .004; however, overestimation on Raven's Progressive Matrices was uncorrelated with overestimation on the weight-guessing task, r = −.01, p = .921. Overplacement on the baseball forecasting task was positively correlated with overplacement on Raven's Progressive Matrices, r = .19, p < .001, and on the weight-guessing task, r = .12, p = .021, but overplacement on Raven's Progressive Matrices was not correlated with overplacement on the weight-guessing task, r = .01, p = .862. We obtained high correlations between overprecision measures for each pair of tasks, rs > .6, ps < .001, all dfs = 379.

[1]Remarkably, the negative numbers mean that our participants were actually *underprecise* in their predictions of the distribution of all participants' scores on all three tasks. This may be consistent with Moore et al. (2015), who find that people's subjective probability distributions tend to be spread too widely (underprecise) but poorly centered, especially on tasks that are more chance-based or less familiar.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 2. *Studies 1a and 1b inter-task correlations for accuracy, confidence, and overconfidence.*

	Study 1a ($N = 379$)			Study 1b ($N = 138$)					
Measure	MLB-RPM	MLB-WGT	RPM-WGT	MLB-RPM	MLB-WGT	RPM-WGT	GOT-MLB	GOT-RPM	GOT-WGT
Accuracy (Score)	.15**	−.03	−.08	.09	.07	.02	−.02	.00	.07
Confidence									
Estimation	.19***	.21***	.09	.18*	.22**	.17	.19*	.18*	.20*
Placement	.21***	.21***	.15**	.17	.24**	.26**	.26**	.17	.30***
Placement (Percentile)	.28***	.31***	.19***	.21*	.35***	.27***	.20*	.30***	.15
SPD Variance	.63***	.73***	.67***	.64***	.64***	.69***	.60***	.60***	.71***
Overconfidence									
Overestimation	.22***	.15**	−.01	.19*	.20*	.13	.08	.20*	.07
Overplacement	.19***	.12*	.01	.20*	.27*	.17*	.22*	.15	.14
Overplacement (Percentile)	.07	.11*	.04	.27*	.16	.28***	.06	.14	.19*
Overprecision	.63***	.73***	.67***	.64***	.64***	.69***	.60***	.60***	.71***

Note: Asterisks denote results of two-tailed tests comparing the correlations to 0, *$p<.05$, **$p<.01$, ***$p<.001$.

Although the focus of this article is on overconfidence rather than confidence, some have argued that subtracting accuracy from confidence may mostly add noise (Binnendyk and Pennycook, 2024). In this study, one could make this argument for both the MLB forecasting task and the weight-guessing task, as confidence (score estimate) was not correlated with accuracy (actual score) on either task, $r_{MLB} = .08$, $p = .125$, $r_{WGT} = -.03$, $p = .596$. Thus, we report correlations between confidence measures in addition to overconfidence measures in Table 2. Strikingly, participants' confidence as measured by their score estimates, score estimates minus estimates of others' average scores, and narrowness of their subjective probability distribution of others' scores, are all positively correlated with each other across tasks (with the exception of the relationship between score estimation on the RPM and weight-guessing tasks, $r(377) = .09$, $p = .078$).

2.2.2. Correlations between overconfidence and trait measures

We report all correlations between overconfidence measures and trait measures in Table 3, and all correlations between trait measures and confidence and accuracy in the Supplementary Material. Actively open-minded thinking was negatively correlated with both overestimation, $r = -.15$, $p = .004$, and overplacement, $r = -.11$, $p = .029$, on Raven's Progressive Matrices. Notably, actively open-minded thinkers were also more accurate on this task, $r = .21$, $p < .001$, but did not estimate their performance significantly more highly, $r = .07$, $p = .196$, or place, $r = .10$, $p = .059$. Surprisingly (to us), actively open-minded thinking was positively correlated with precision on all three tasks, significantly on the MLB forecasting and weight-guessing tasks and directionally on Raven's Progressive Matrices. In other words, actively open-minded thinkers had narrower subjective probability distributions; given that our participants were on average underprecise with regard to the score distribution of all participants, actively open-minded thinkers were actually better calibrated.

Male gender correlated with *over*placement only on the weight-guessing task, $r = .14$, $p = .007$. However, gender was actually correlated with placement on all three tasks, $r_{MLB} = .24$, $p < .001$, $r_{RPM} = .17$, $p < .001$, $r_{WGT} = .13$, $p = .012$; in other words, men thought they would perform better than other participants in the study on all three tasks, compared to women. However, it seems that being male was actually correlated with better performance on both the MLB forecasting task, $r = .17$, $p = .001$, and Raven's Progressive Matrices, $r = .16$, $p = .002$, so only overplacement on the weight-guessing task remains significantly positive after subtracting performance.

Finally, we note that age was positively correlated with overestimation on the weight-guessing task, $r = .20$, $p < .001$. Interestingly, this is driven by both a positive relationship between age and score estimation, $r = .13$, $p = .013$, and a negative relationship with accuracy, $r = -.18$, $p < .001$. We do not have strong hypotheses about why age would increase overestimation on this task alone nor does this relationship replicate in Study 1b.

2.3. Discussion

Our observation that overprecision correlates across tasks is consistent with previous literature; for example, Moore and Healy (2008) report an alpha of 0.95 when looking at overprecision across a set of trivia tests. Our inter-task correlations for overestimation and overplacement vary between .03 and .25, similar to the range West and Stanovich (1997) report and lower than the .50 reported by Klayman et al. (1999). While some of these inter-task correlations are positive, in absolute terms it seems that only our measurement of overprecision is consistently correlated between tasks. We acknowledge that all of the correlations are positive; we interpret this as weak (but nonzero) support for within-individual stability of overconfidence as a trait. Further, we note that the inter-task correlations between confidence measures are all positive. However, because we are studying *over*confidence, we believe that subtracting accuracy or otherwise adjusting for it is necessary by definition.

Interestingly, we do observe some correlations between trait measures and overconfidence, though they appeared to be task-specific. Actively open-minded thinkers overestimated themselves less on Raven's Progressive Matrices, which is a task where it is relatively easier to know when you have

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 3. *Studies 1a and 2: Correlations between overconfidence and trait measures.*

Overestimation	AOT	Narcissism	OCT	Gender (M)	Age
[1a] MLB	−.07	.01		.00	.07
[1a] RPM	−.15***	−.05		−.07	.06
[1a] WGT	.00	.00		.06	.20***
[1b] MLB	.05	−.07		.05	.07
[1b] RPM	−.12	−.01		.00	.09
[1b] WGT	.16	.04		.28***	.10
[1b] GOT	−.03	−.07		.30***	−.02
[2] NFL (T1)	−.10	−.04	.21*	−.02	−.02
[2] NFL (T2)	.01	.02	.03	.13	−.10
Overplacement	**AOT**	**Narcissism**	**OCT**	**Gender (M)**	**Age**
[1a] MLB	.03	−.07		.07	−.04
[1a] RPM	−.11*	−.09		.01	.01
[1a] WGT	.06	−.04		.14**	.16**
[1b] MLB	−.03	−.11		.12	−.03
[1b] RPM	−.10	−.13		.12	.09
[1b] WGT	.13	.01		.32***	.04
[1b] GOT	−.03	−.10		.22*	−.14
[2] NFL (T1)	−.06	−0.08	.11	.03	−.02
[2] NFL (T2)	−.01	.13	−.06	.23**	−.05
Overprecision	**AOT**	**Narcissism**	**OCT**	**Gender (M)**	**Age**
[1a] MLB	.18***	.12*		−.01	.05
[1a] RPM	.10	.07		−.05	−.01
[1a] WGT	.16*	.09		−.05	.02
[1b] MLB	−.03	−.09		.14	.00
[1b] RPM	.01	−.13		.03	.02
[1b] WGT	.09	−.09		.04	.02
[1b] GOT	.08	−.10		.13	0.00
[2] NFL (T1)	.18*	.13	−.04	.07	.01
[2] NFL (T2)	0.16	0.12	−0.10	0.03	0.08

Note: Asterisks denote results of two-tailed tests comparing the correlations to 0, *p<.05, **p<.01, ***p<.001.

arrived at the correct answer. Men placed themselves higher than women on all three tasks, though they only overplaced more on the weight-guessing task as they performed better on the MLB forecasting and Raven's Progressive Matrices task.

3. Study 1b: Overconfidence across domains and time

In Study 1b, we followed up with the participants from Study 1a after 10 months with a nearly identical survey; the only difference was that we added a fourth task to elicit overconfidence measures—the Generalized Overconfidence Task from Binnendyk and Pennycook (2024). The two primary purposes of this study were to replicate findings from Study 1a on weak cross-domain relationships and relationships with other trait measures and determine test-retest reliability for the three measures of overconfidence on Raven's Progressive Matrices and the weight-guessing task.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

3.1. Method

We preregistered this study at https://aspredicted.org/MNZ_W2V on May 21, before data collection began on May 22, 2024.

3.1.1. Participants

We recruited as many of our original Study 1a participants as possible before the forecasted events (professional baseball games) passed. 151 of our original 381 participants entered the survey. We excluded five participants who straightlined the actively open-minded thinking scale, one who straightlined the narcissism scale, two participants whose reported estimates for the average of other scores' differed by more than two from the mean of their reported subjective probability distributions for the same estimate on all four tasks, and five participants who failed one of our two bogus-item attention checks, leaving us with a final sample of 138 participants who had usable data at both time points. Our participants were 49.69% male, 49.07% female, 1.24% other gender; 68.32% White / 4.35% Asian / 14.29% Black / 5.59% Hispanic / 7.45% Other, M_{age} = 38.45, SD_{age} = 11.36. Participants earned $6.00 for participating in the study, which on average took 27.91 min.

We compared those who did and did not return on the following data from Study 1a: demographics (age, gender), actively open-minded thinking, and each measure of accuracy and confidence. We note that the 153 participants who returned for Study 1b were significantly older than those who did not return (M_{1b} = 38.07, SD = 11.76; $M_{1a\ only}$ = 33.51, SD = 9.27), $t(234.18) = 3.92$, $p < .001$, $d = 0.45$, and overplaced less on Raven's Progressive Matrices than those who did not return ($M_{1b} = -0.49$, $SD = 2.25$; $M_{1a\ only} = 0$, $SD = 2.24$), $t(283.65) = 2.07$, $p = .039$, $d = 0.22$. These two groups of participants did not significantly differ on any other demographics, confidence, or overconfidence measures (see the Supplementary Material for details).

3.1.2. Procedure

Study 1b was similar to Study 1a, with the following changes. For the MLB forecasting task, participants predicted the outcomes of 10 MLB games that occurred on June 2, 2024. We added a fourth task, Binnendyk and Pennycook's (2024) Generalized Overconfidence Task, in which participants try to guess which of two choices is pictured in a fuzzy image. Participants see 10 nearly indiscernible blurred images, each for 0.25 s, and guess whether the image is one of two options (e.g., a chimpanzee or a baseball player). Participants answered the same set of 10 questions for the Raven's Progressive Matrices and saw the same 10 images for the weight guessing task.

3.1.3. Measures

Actual Scores (Accuracy). As in Study 1a, we measured accuracy the same way in Study 1a (M_{MLB} = 4.85, SD = 1.41; M_{RPM} = 4.29, SD = 2.04; M_{WGT} = 1.59, SD = 1.48; M_{GOT} = 5.57, SD = 1.61).

Confidence (Estimation). As in Study 1a, we measured estimation as participants' estimates of their own scores (M_{MLB} = 5.08, SD = 1.77; M_{RPM} = 3.99, SD = 2.09; M_{WGT} = 5.18, SD = 1.84; M_{GOT} = 2.36, SD = 1.97).

Confidence (Placement). As in Study 1a, we measured placement primarily by asking participants to estimate the average score of all participants in the study (M_{MLB} = 5.22, SD = 1.27; M_{RPM} = 4.65, SD = 1.40; M_{WGT} = 4.95, SD = 1.34; M_{GOT} = 3.94, SD = 1.67) and then subtracting participants' estimates of others' average score from their estimates of their own score ($M_{MLB} = -0.14$, SD = 1.72; $M_{RPM} = -0.66$, SD = 2.04; M_{WGT} = 0.23, SD = 1.54; $M_{GOT} = -1.61$, SD = 2.20), and secondarily by measuring their estimate of their score's percentile (M_{MLB} = 38.29, SD = 23.20; M_{RPM} = 35.32, SD = 24.46; M_{WGT} = 37.55, SD = 22.70; M_{GOT} = 29.15, SD = 27.93). These two measures of placement were correlated with each other on all three tasks, $r_{MLB} = .46$, $r_{RPM} = .53$, $r_{WGT} = .45$, $ps < .001$.

Confidence (Precision). As in Study 1a, we measured precision as the variance of participants' subjective probability distributions (M_{MLB} = 5.23, SD = 2.47; M_{RPM} = 5.23, SD = 2.70; M_{WGT} = 5.47, SD = 2.61; M_{GOT} = 5.16, SD = 2.94).

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 4. *Study 1a–1b test–retest reliability for accuracy, confidence, and overconfidence.*

Measure	MLB	RPM	WGT
Accuracy (Score)	.11	.55***	.37***
Confidence			
Estimation	.35***	.50***	.44***
Placement	.30***	.39***	.48***
Placement (Percentile)	.20*	.18*	.27***
SPD Variance	.43***	.36***	.46***
Overconfidence			
Overestimation	.19*	.33***	.42***
Overplacement	.17*	.21*	.46***
Overplacement (Percentile)	.06	.35***	.25**
Overprecision	.43***	.36***	.46***

Overestimation. As in Study 1a, we measured overestimation by subtracting actual from estimated scores (M_{MLB} = 0.23, SD = 2.16; M_{RPM} = −0.30, SD = 2.06; M_{WGT} = 3.59, SD = 2.53; M_{GOT} = −3.21, SD = 2.53).

Overplacement. As in Study 1a, we calculated overplacement primarily by subtracting other-placement from self-placement (M_{MLB} = −0.16, SD = 2.17; M_{RPM} = −0.80, SD = 1.95; M_{WGT} = 0.22, SD = 2.26; M_{GOT} = −12.77, SD = 2.81), and secondarily by subtracting actual percentile rank from estimated percentile rank (M_{MLB} = −2.60, SD = 35.61; M_{RPM} = −10.00, SD = 30.85; M_{WGT} = −3.68, SD = 39.79; M_{GOT} = −11.43, SD = 38.24).

Overprecision. As in Study 1a, we measured overprecision by subtracting the variance of participant's subjective probability distributions of others' scores from the actual variance (M_{MLB} = −3.12, SD = 2.47; M_{RPM} = −0.92, SD = 2.70; M_{WGT} = −3.33, SD = 2.61; M_{GOT} = −2.55, SD = 2.94).

Narcissism. We used the same scale for narcissism as in Study 1a (α = 0.75, M = 12.32, SD = 3.10), except that we added one 'bogus' measure to serve as an attention check ('I am paid bi-weekly by leprechauns') which was not included in the scale calculation.

Actively open-minded thinking. We used the same scale for AOT as in Study 1a (α = .72, M = 3.79, SD = 0.54), except that we added one 'bogus' measure to serve as an attention check ('I have had a fatal heart attack') which was not included in the scale calculation.

3.2. Results

3.2.1. Correlations across time

First, we calculated cross-time correlations for each measure of accuracy, confidence, and overconfidence that were in both Studies 1a and 1b (Table 4). All nine measures of overconfidence (three measures on three tasks) were significantly positively correlated between time points; in general, cross-time correlations were higher on Raven's Progressive Matrices and the weight-guessing task (perhaps unsurprisingly since they were the exact same questions).

Accuracy was correlated across time points on both Raven's Progressive Matrices, r = .55, p < .001, and the weight-guessing task, r = .37, p < .001, but not on the MLB forecasting task, r = .11, p = .209. With the exception of placement on the MLB forecasting task, all measures of confidence were also positively correlated with themselves: score estimation, r_{MLB} = .35, r_{RPM} = .50, r_{WGT} = .44, ps < .001; the difference between self-score estimation and the average of others' score estimation, r_{MLB} = .05, p = .535, r_{RPM} = .30, p < .001, r_{WGT} = .17, p = .046, percentile estimates of scores, $\mathrm{r}_{\mathrm{MLB}}$ = .30, r_{RPM} = .39, r_{WGT} = .48, ps < .001, and variance on subjective probability distributions of other's

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

scores (same correlations as inter-time correlations of overprecision, since we are simply subtracting constants).

3.2.2. Correlations across domains

We report all correlations between tasks for confidence, accuracy, and overconfidence in Table 2. In general, we observe similar patterns and magnitudes of results as in Study 1a. Average inter-task correlations for each form of overconfidence are $r = .15$ for overestimation, $r = .19$ for overplacement, and $r = .65$ for overprecision. In fact, the inter-task correlations for overconfidence were *higher* than in Study 1a. Overestimation was again significantly positively correlated between the MLB and RPM tasks, $r = .19$, $p = .026$, and the MLB and weight-guessing tasks, $r = .20$, $p = .017$. Overplacement was positively correlated between all pairs of tasks (insignificantly between the GOT and RPM and the GOT and WGT, but we note we may be underpowered in Study 1b to detect such correlations). As in Study 1a, inter-task correlations between overprecisions, all rs > .60. We also note that inter-task correlations for each confidence measure are all positive, rs > .15, though accuracy was not correlated between any of the six pairs of tasks.

3.2.3. Correlations with trait measures

We report all correlations between overconfidence and trait measures[2] in Table 3. For the most part, we did not find statistically significant replications of most of the relationships that we observed in Study 1a. However, we note that we have significantly less power to detect these relationships than in Study 1a and thus we should expect statistically weaker results. Interestingly, we did find some differences by gender; men overestimated more on both the weight-guessing task, $r = .30$, and the GOT, $r = .28$, ps < .001. They also overplaced more on the same two tasks, $r_{WGT} = .32$, $p < .001$, $r_{GOT} = .22$, $p = .011$. This is driven by men estimating that they would score more highly on these two tasks, $r_{WGT} = .19$, $p = .025$, $r_{GOT} = .21$, $p = .015$, but in reality scoring worse, $r_{WGT} = -.24$, $p = .005$, $r_{GOT} = -.21$, $p = .015$. As in Study 1a, men placed themselves relatively higher than women did on the MLB task, $r = .22$, the RPM task, $r = .24$, and the weight-guessing task, $r = .24$, however, this only manifested into greater overplacement on the weight-guessing task.

3.3. Discussion

Study 1b seems to provide evidence for the reliability of the measures across time on accuracy, confidence, and overconfidence. As expected, we observe stronger inter-task consistency for overprecision than for overestimation and overplacement. All nine overconfidence measures in Studies 1a and 1b were positively correlated with themselves after 10 months (rs 0.17–0.46), though their inter-time correlations were weaker than other measures (e.g., the inter-time correlation for AOT was $r = 0.71$). When examining inter-task correlations using only data from Study 1a, we replicate the patterns of results from Study 1a demonstrating positive but weak correlations between task domains for overestimation and overplacement, and stronger positive correlations between overprecision on different task domains as measured by subjective probability distribution spread. Interestingly, we did replicate results from Study 1a suggesting that men place (and perhaps) overplace themselves higher than women.

[2]The correlations we report here use the measurements of actively open-minded thinking and narcissism from Time 2 (Study 1b) rather than Time 1 (Study 1a), and the demographics we use (gender, age) are the ones collected at Time 1 (Study 1a). Participants's AOT scores were very consistent between time points, $r(159) = .72$, as were their narcissism scores, $r(159) = .73$, ps < .001.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

4. Study 2: Overconfidence across time

Study 2 examines whether overconfidence measures are consistent across time on a smaller scale than Study 1a; we use a single task (professional football forecasting) over the course of 1 week. However, we told participants that there would be a second part of the study so there would be significantly less attrition than in Study 1b. We also derived an exploratory measure of overestimation specific to the context (sports fans making predictions about their teams).

4.1. Method

We preregistered this study at https://aspredicted.org/HLV_TB8 on November 1, 2023 before data collection started on November 2, 2023.

4.1.1. Participants

We initially distributed the survey to 257 English-speaking CloudResearch participants located in the US, with an approval rating >95% and <1000 HITs. Participants then answered three screening questions; the first question asked whether they followed the National Football League (NFL), and the next two questions asked about the results of the most recent NFL Championship. Participants who answered 'No' to the first question or answered any of the other two questions incorrectly could not proceed with the survey. 73 participants failed these screening criteria, an additional 12 began the survey after some of the forecasted games had started, and we excluded one participant who straightlined the 16-item narcissism scale, leaving us with 171 participants at Time 1. 36 of these participants did not complete the follow-up survey, leaving us with a final sample of 135. Our participants were 42.69% male, 69.01% White / 4.09% Asian / 11.7% Black / 5.85% Hispanic / 9.36% other race; M_{age} = 35.93, SD_{age} = 9.77; with a median education level of an Associate's degree in college. Participants earned $1.20 for participating in the first study, which on average took 10.69 min, and $1.50 for participating in the second part of the study, which on average took 6.05 min.

While participants who did and did not return did not significantly differ on gender, age, narcissism, accuracy, or any of our primary confidence and overconfidence measures (see the Supplementary Material for details), the 36 participants who did not return for the follow-up survey picked their favorite teams as winners marginally more often (M = 1.61, SD = 2.05) than those who did return (M = 0.98, SD = 1.00), $t(39.52) = 1.80$, $p = .079$, $d = .49$, and overestimated the game wins of their favorite teams ($M = 0.5$, SD = 0.81) more than those who did return (M = 0.2, SD = 0.73), $t(51.19) = 2.01$, $p = .049$, $d = 0.40$. In addition, the participants who returned were marginally more precise—that is, had less variance—in their subjective probability distributions of the overall score distribution (M = 5.22, SD = 2.15) than those who did return (M = 5.91, SD = 2.16), $t(54.86) = 1.70$, $p = .096$, $d = 0.32$.

4.1.2. Procedure

We distributed the two surveys 1 week apart. For each survey, participants guessed the winners of 10 NFL games that would take place the following Sunday. They then provided the following confidence measures (analogous to those in Studies 1a and 1b): an estimate of how many game winners they predicted correctly, an estimate of the average number of game winners other participants predicted correctly, an estimate of their own score's percentile, and a subjective probability distribution for the distribution of all participants in the study.

In the first survey, participants then indicated their favorite NFL teams ('Which of the following NFL Teams do you consider yourself a fan of? (Mark all that apply)') and completed the following trait measures: narcissism (same scale as Studies 1a and 1b), four items from the actively open-minded thinking (AOT) scale, and Overconfidence Test (Lawson et al., 2023). Lastly, they reported their demographics.

In the second survey, participants then completed the remaining three items of the actively open-minded thinking (AOT) scale that they did not complete in the first survey.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

4.1.3. Measures

Actual Score (Accuracy). As in previous studies, we calculated participants' scores (out of 10) as the number of game winners forecasted correctly. Performance was significantly above chance at Time 1 (M_{T1} = 5.67, SD_{T1} = 1.57), $t(134) = 4.98$, $p < 0.001$, $d = 0.43$, but not at Time 2 (M_{T2} = 4.96, SD_{T2} = 1.47).

Confidence (Estimation). As in previous studies, participants estimated their scores out of 10 (M_{T1} = 6.54, SD = 1.49; M_{T2} = 5.79, SD = 1.69). We also constructed an exploratory measure of estimation by counting the number of times participants predicted that one of their favorite teams would win (M_{T1} = 0.98, SD = 1.00; M_{T2} = 0.86, SD = 0.84).

Confidence (Placement). As in previous studies, participants estimated the average score of all participants in the study (M_{T1} = 6.02, SD = 1.31; M_{T2} = 5.54, SD = 1.33), which we subtracted from their estimates of their own scores for a measure of placement (M_{T1} = 0.52, SD = 1.97; M_{T2} = 0.25, SD = 1.82). Our secondary measure of placement asked participants to estimate their percentile rank among participants in the study (M_{T1} = 49.79, SD = 20.11; M_{T2} = 43.40, SD = 19.30). These two measures of placement correlated with each other, $r_{T1} = .35$, $r_{T2} = .33$, $ps < .001$.

Confidence (Precision). As in previous studies, we calculated the variance of participants' subjective probability distributions of the scores of all participants in the study (M_{T1} = 5.22, SD = 2.15; M_{T2} = 4.91, SD = 1.99).

Overestimation. As in previous studies, we calculated overestimation by subtracting participants' actual scores from their estimated scores (M_{T1} = 0.87, SD = 2.00; M_{T2} = 0.83, SD = 2.22). We also constructed an exploratory measure of overestimation only using predictions about games where participants indicated they were a fan of one of the teams playing; we counted the number of times participants predicted that one of their favorite teams would win and subtracted the number of times one of their favorite teams actually won.

Overplacement. As in previous studies, we calculated overplacement as (estimated own score − estimate of mean score of all participants) − (actual own score − actual mean score of all participants) (M_{T1} = 0.52, SD = 2.30, M_{T2} = 0.25, SD = 2.30). Our secondary measure of overplacement subtracted participants' actual percentile rank from their estimated percentile rank (M_{T1} = 8.49, SD = 33.14; M_{T2} = 3.19, SD = 33.30). These two measures of overplacement correlated with each other, $r_{T1} = .60$, $r_{T2} = .62$, $ps < .001$.

Overprecision. As in previous studies, we calculated overprecision by subtracting the variance of participants' subjective probability distributions of all participants' scores from the true variance of all participants' scores ($M_{T1} = -2.74$, SD = 2.15; $M_{T2} = -2.75$, SD = 1.99).

Actively Open-Minded Thinking (AOT): We used the same scale as in prior studies ($\alpha = 0.75$, M = 3.29, SD = 0.43). However, we split this scale into two parts with four items being answered at Time 1 and the remaining three at Time 2.

Narcissism (NPI): We used the same scale as in prior studies ($\alpha = 0.65$, M = 12.99, SD = 2.44).

Overconfidence Test (OCT): We used the 3-item Overconfidence Test (OCT) (Lawson et al., 2023). Each item ranges from 0 to 100%, e.g., 'One hundred people are guessing the number of jellybeans in a jar. The closest 10 guesses win $100. How likely are you to be one of the winners?', $\alpha = 0.45$, M = 33.81, SD = 13.42).

4.2. Results

4.2.1. Correlations between time points

All three inter-time correlations between the primary overconfidence measures were positive and significant: $r_{overestimation} = 0.29$, $p < .001$; $r_{overplacement} = .21$, $p = .015$; $r_{overprecision} = .49$, $p < .001$. The inter-time correlation for the exploratory measure of overconfidence based on whether participants thought their favorite teams would win was also positive, $r = .22$, $p = .010$. The secondary percentile-based measure of overplacement was not significantly correlated between time points, $r = 10$, $p = .227$.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

These positive correlations were driven more by inter-time correlations in confidence than accuracy. Participants' scores at time 1 were not significantly correlated with their scores at Time 2, $r = .14$, $p = .117$. However, their confidence was correlated on all three primary and both secondary measures of confidence: self-score estimations, $r = .46$, the number of favorite teams they predicted winning, $r = .39$, the difference between self-score estimates and all-participant-average estimates, $r = .39$, estimates of own score percentiles, $r = .35$, and subjective probability distribution of all participants' scores, $r = .49$, $ps < .001$.

4.2.2. Correlations between overconfidence measures and trait measures

Table 3 shows correlations between the three primary overconfidence measures at both time points and the stable trait measures (actively open-minded thinking, narcissism, Overconfidence Test, gender, and age). Participants who scored higher on the overconfidence test were also more likely to overestimate their scores at Time 1, $r = .21$, $p = .016$, though not at Time 2, $r = .03$, $p = .705$. Participants who scored higher on actively open-minded thinking seemed to again have narrower confidence intervals at Time 1, $r = .18$, $p = .032$, and directionally at Time 2, $r = .16$, $p = .061$.

We also found that males were more likely to overplace at Time 2, $r = .23$, $p = .009$, but not at Time 1, $r = .03$, $p = .69$. Digging into the precise components of overplacement, we find that males estimated that they would score higher than non-males at both Time 1, $r = .20$, and Time 2, $r = .31$, and that they would score higher relative to other participants in the study (estimated self-score minus estimated all-participant-score) at Time 1, $r = .22$, and even more so at Time 2, $r = .42$. They did in fact score higher than non-males at both Time 1, $r = .22$, and Time 2, $r = .17$.

Interestingly, we observed multiple statistically significant correlations between the exploratory measure of overestimation based on participants' favorite teams at Time 1. Participants who overestimated the performance of their favorite teams scored lower on actively open-minded thinking, $r = -.20$, $p = .07$, higher on the Overconfidence Test, $r = .23$, $p = .008$ and were younger, $r = -.22$, $p = .011$. None of these correlations showed up at Time 2, but we then observed that men overestimated their favorite teams less, $r = -.24$, $p = .006$. We find these results interesting but note that because the participants who most overestimated their favorite teams were also more likely to attrit, we interpret these results with caution.

4.3. Discussion

Overconfidence measures seemed durable across time within individuals. As one might expect given the shorter timespan (1 week vs. 10 months), these correlations were stronger than those in Studies 1a and 1b on the most analogous task, the baseball forecasting task. The positive correlations seemed to be driven by consistency in confidence rather than accuracy. Still, the strength of correlations we observed between time points for overestimation ($r = .29$) and measures of overplacement ($r = .22$) differ dramatically from other personality measures taken a week apart, such as the Big-Five: extraversion ($r = .92$), agreeableness ($r = .92$), conscientiousness ($r = .92$), openness ($r = .93$), and emotional stability ($r = .91$) (Kurtz and Parrish, 2001), or from our measures of AOT ($r = .71$) or narcissism ($r = .73$) 10 months apart in Studies 1a and 1b. Although overestimation and overplacement demonstrated some consistency across time, we interpret the strengths of these correlations as too low to qualify as stable traits. By contrast, overprecision correlated across time points at $r = .49$, again demonstrating that the most stable form of overconfidence of our current measures is overprecision.

Interestingly, we did observe positive correlations between overprecision and actively open-minded thinking, which is directionally consistent with Study 1a. As in Study 1a's sports forecasting task, we found that males were both more confident and more accurate in their forecasts. Although we did observe some statistically significant correlations between the overconfidence measures and trait measures, we note that they were inconsistent between the two-time points in our data and that some could be spurious given a large number of correlation tests.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

5. Study 3a: Multiple measures of precision

Study 3a examines the relationship between different measures of confidence—specifically precision, or certainty that a belief is correct—on the same Generalized Overconfidence Task (Binnendyk and Pennycook, 2024) from Study 1b. Given that our prior studies point towards overprecision potentially being the most reliable form of overconfidence, we wanted to test how reliably different measures of precision—and presumably overprecision—correlated with each other. We selected a wide range of precision measures, with the intent of testing the degree to which different elicitations of 'How certain are you?' correlate with each other; we selected classic measures of Likert scales and confidence interval widths, coupled with the more modern subjective probability distribution and an incentive-compatible bet. In addition, this study employs a measure of precision that is directly related to participants' predictions (estimates of their own scores), rather than a second-order belief about other participants' scores.

The studies we have reported thus far found some mixed and ambiguous evidence for trait-like consistency. Study 3a sought to establish clearer benchmarks by which we can assess the consistency of these precision measures; we attempt to ask not only the degree to which different measures of precision correlate with each other, but also whether these correlations are lower than one might expect or lower than they should be if they were tapping the same underlying trait, even accounting for noisiness in the data. In other words, we ask whether participants' reports of precision seem to vary more as a function of the elicitation method or situation than as a function of some consistent internal trait. With this goal in mind, we collected two benchmarks: expert predictions and simulated data. We provide more detail on both of these below.

Finally, Studies 3a and 3b employ a wider range of additional trait measures than Studies 1a–2; in addition to actively open-minded thinking and the overconfidence test from Study 2, we also test whether certainty measures correlate with factors of intellectual humility.

5.1. Method

We preregistered this study at https://aspredicted.org/blind.php?x=Z29_FCB on October 3, 2023 before data collection started on October 4, 2023.

5.1.1. Participants

We recruited 326 participants from Cloud Research. We excluded 66 participants who answered fewer than three out of five comprehension questions about subjective probability distributions correctly (see below), 13 additional participants whose 90% confidence interval endpoints were out of order (i.e., lower bound higher than upper bound), two participants who straightlined the actively open-minded thinking scale, and three participants for whom both of their reported score estimates (for self and average of all participants) was more than two away from the mean of the corresponding subjective probability distribution. This left us with a final sample of 242. Our participants were 32.92% male, 67.9% White / 2.88% Asian / 11.11% Black / 3.7% Hispanic / 14.4% Other, M_{age} = 35.84, SD_{age} = 9.78, with median education level a Bachelor's degree in college. Participants earned $1.45 in addition to a $1.00 bonus that they could choose to bet. The median completion time was 14.22 min.

5.1.2. Procedure

Participants learned that they were going to 'play a game where they identify the objects in a scrambled image (10 images total),' and then completed the Generalized Overconfidence Task. On average, participants identified 6.53 (SD = 1.50) out of 10 images correctly.

Afterward, participants answered a five-question comprehension check (M = 3.82 among participants we included, with 41 participants answering all five correctly), assessing their understanding of subjective probability distributions. They were asked what the most likely score would be if someone guessed for all 10 questions in the image task (answer: 5/10), what the summed probability across bins in the subjective probability distribution should equal (answer: 100%), and three multiple-choice

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

questions where they identified the pictured subjective probability distribution that best fit a belief written in words (e.g., 'I think that I correctly guessed the objects in eight out of the 10 images. Which of the tables below best captures my beliefs?'). After each of the five questions, participants learned which answer was correct along with an explanation.

Participants then estimated their own score out of 10 on the task ($M = 4.39$, $SD = 1.29$), and reported their confidence in that prediction with each of the following elicitation methods in a randomized order: a Likert scale, an incentivized bet, a 90% Confidence Interval, and a Subjective Probability Distribution. See 'Certainty Measures' below for more details. Participants also reported their predictions about the performance of other participants in the study as a point estimate ($M = 4.53$, $SD = 1.08$) and in a Subjective Probability Distribution. See 'Other Overconfidence Measures' below.

Finally, participants completed the following trait measures: actively open-minded thinking (Haran et al., 2013), intellectual humility—Factor 1, independence of ego and intellect, intellectual humility—Factor 4, lack of intellectual overconfidence (Krumrei-Mancuso and Rouse, 2016), the Overconfidence Test (Lawson et al., 2023), and demographic measures (gender, age, and race).

5.1.3. Measures

Likert. Participants indicated their certainty that their actual score was within one point of their estimate (above or below) with a 7-point scale, from 1 = 'Not confident at all' to 7 = 'Certain' ($M = 3.59$, $SD = 1.50$). Greater confidence scores correspond to higher precision.

Bet. Participants learned that they would receive a $1.00 bonus, which they could choose to keep or bet any amount up to the full $1 on the accuracy of their estimate ($M = \$0.39$, $SD = \$0.35$). If their estimate was within 1 point of their actual score, the money they bet was doubled, otherwise it was lost. The more money a participant bet on the accuracy of their estimate, the higher their implicit certainty. About 56 participants chose to bet nothing, 40 chose to bet the full $1.00, 47 chose to bet $0.50, and the others bet something else.

90% Confidence Interval. Participants indicated endpoints of a 90% confidence interval; they were instructed to 'identify two numbers: one BELOW your estimate and another ABOVE your estimate. These numbers should be far enough apart that you are 90% sure your true score is between them'. We reverse-scored interval widths ($M = -5.01$, $SD = 2.30$) so that higher numbers, that is, narrower intervals, would correspond to higher certainty.

Subjective Probability Distribution (own scores). Participants reported a probability, from 0 to 100, for each of 11 mutually exclusive and exhaustive bins (e.g., 0/10, 1/10,. . . 10/10). For analysis, we calculated two measures of precision from these subjective probability distributions. First, we calculated peak probability ($M = 0.54$, $SD = 0.18$) by normalizing the probability distribution, finding the leftmost bin with the highest probability assigned, and summing the probability of that bin, the bin immediately below, and the bin immediately above. Second, we calculated variance ($M = 4.86$, $SD = 2.49$) as in prior studies.

We also calculated the same measures of overestimation ($M = -2.14$, $SD = 1.89$), indirect overplacement based on estimating the average of all participants' scores ($M = -0.15$, $SD = 2.02$), and overprecision ($M = -3.32$, $SD = 2.98$) as in prior studies.

Trait Measures. We collected the following trait measures.

Actively open-minded thinking. We used the same scale as in prior studies, $\alpha = 0.76$, $M = 4.15$, $SD = 0.47$.

Intellectual humility, independence of intellect, and ego factor. We used the first factor of the Comprehensive Intellectual Humility Scale: independence of intellect and ego (Krumrei-Mancuso and Rouse, 2016), five items, for example, 'When someone disagrees with ideas that important to me, it feels as though I'm being attacked' (reverse-scored) from 1 = 'strongly disagree' to 5 = 'strongly agree', $\alpha = 0.91$, $M = 3.45$, $SD = 0.94$.

Intellectual humility, lack of intellectual overconfidence. We used the fourth factor of the Comprehensive Intellectual Humility Scale: lack of intellectual overconfidence (Krumrei-Mancuso and Rouse, 2016), five items, for example, 'I feel small when others disagree with me on topics that are close to

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 5. *Study 3a: Correlation tests comparing observed correlations to rational simulated benchmarks and expert-predicted benchmarks.*

			Correlation		Simulated		SSPP		
Measure 1	Measure 2	Observed	Benchmark (Simulated)	Benchmark (SSPP)	$t_{rational}$	$p_{rational}$	t_{sspp}	p_{sspp}	n_{sspp}
Bet	CI Width	−.01	.43	.27	7.36	<.001	4.49	<.001	24
Bet	SPD Peak	.07	.45	.38	6.29	<.001	5.14	<.001	27
CI width	SPD Peak	.10	.49	.45	6.65	<.001	2.34	.020	24
Likert	Bet	.31	.44	.42	2.45	.015	2.05	.041	23
Likert	CI Width	.03	.48	.30	7.71	<0.001	4.36	<.001	23
Likert	SPD Peak	.17	.49	.34	5.66	<0.001	2.90	.004	24

Note: Correlations between pairs of confidence measures, compared to benchmarks via a Fisher's *z*-transformation. Expert predictors from SSPP were only required to predict the correlation between Bet and Subjective Probability Distribution Peak (SPD Peak), and were not required to predict the other five pairwise correlations. Confidence Interval Width is reverse-scored.

my heart' (reverse-scored) from 1 = 'strongly disagree' to 5 = 'strongly agree', $\alpha = 0.75$, $M = 3.40$, $SD = 0.61$. We note that the items for intellectual humility on both factors most often refer to reactions or comparisons to other people, and would thus likely be most related to overplacement if it were related to any of the three forms.

Overconfidence Test (OCT): We used the same scale as in Study 2, $\alpha = 0.42$, $M = 31.43$, $SD = 13.38$.

5.1.4. Benchmarks

We compare each pairwise correlation between confidence measures to two preregistered benchmarks.

Expert Predictions. We asked 27 attendees at an October 13, 2023 conference on the Social Science Prediction Platform (SSPP) to make predictions for the strengths of our correlations between precision measures. The respondents mainly consisted of academics, ranging from pre-doctoral research associates to tenured professors, in fields including economics, information science, and psychology. The 27 attendees learned the definition of overprecision, and saw an example of one of the 10 questions from the image-guessing task in Study 3a. Then they learned about four key precision measures—Subjective Probability Distribution Peak, Likert, Bet, and Confidence Interval Width—including the exact wording of how each one would be elicited and coded for analysis. Then, respondents predicted the correlation between Bet confidence and Subjective Probability Distribution Peak confidence. Finally, they were invited (but not required) to make predictions about the other five pairwise correlations. Table 5 shows the averaged predictions, along with the number of people who made each prediction; most of them made all six predictions.

Simulated Benchmarks. We also simulated data to compute benchmarks for what each of the six correlations between the four confidence measures *should* be for a rational agent. We represented each agent i as follows. An agent's belief A about their performance was modeled as a truncated normal distribution with mean M—a random variable drawn uniformly from the interval [0, 10]—and standard deviation S—a random variable drawn uniformly from the interval [0, 10]. The resulting distribution was truncated within the interval [0, 10]. We represented the means and standard deviations of agent's beliefs as random variables to represent participants with many different beliefs—those who believed they did better or worse, and those who were more or less confident in their predictions. A normal distribution reasonably captures the general shape of a belief distribution that is centered at some value and monotonically decreasing on either side; in a truncated normal distribution, the probability mass between the interval (in this case, 0–10) is scaled up to take into account probability mass that would fall outside the bounds of the interval in a normal distribution (Johnson et al., 1995).

$$\mathrm{Ai} \sim \mathrm{TN}\left(\mathrm{M}, \mathrm{S2}, 0, 10\right), \text{where } \mathrm{M} \sim \mathrm{Unif}\left(0, 10\right) \text{ and } \mathrm{S} \sim \mathrm{Unif}\left(0, 5\right)$$

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

We calculate each precision measure as follows. We abbreviate the cumulative distribution function of this normal distribution as CDF. Let p_i be an agent's subjective belief that they are correct (within 1 point of their estimated score).

pi = CDF(M + 1) − CDF(M − 1)
Likert = 1 + pi * (7 − 1)
Bet = 1 if pi > .5, 0 if pi < .5
Peak Bin$i$ = CDF(ceiling(M) + 1) − CDF(floor(M) − 1)
CI Width = $CDF^{-1}(95) - CDF^{-1}(5)$

We calculate p_i by finding the probability mass within the mean +1 and the mean −1. Likert confidence is a simple rescaling of probabilistic confidence, where 1 on the Likert scale corresponds to 0% probability and 7 on the Likert scale corresponds to 100%. For Bet, we assume that a rational agent bets to maximize their expected value; this is 1 if p_i > .5 and 0 if p_i < .5[3]. Peak Bin corresponds to the probability assigned to the mean's bin (where each bin represents a possible score between 0 and 10) and the two neighboring bins; we can calculate this by finding the probability mass between the score below the integer floor of the mean and the score above the integer ceiling of the mean. For a 90% confidence interval, we calculated the distance between the 95th percentile and the 5th percentile of the distribution. We then injected noise into these measures by adding a normally distributed random variable centered at 0 with standard deviation equal to the standard deviation in the data for that measure of precision[4]. Readers may find the code for this simulation in our online repository: https://osf.io/tb2me/files/osfstorage?view_only=77b881f0b92647689270ed3b485366f9.

5.2. Results

5.2.1. *Correlations between confidence measures.*

Table 5 shows the results of each test comparing our observed correlations to each of the two preregistered benchmarks using correlation significance testing and Fisher's z-transformation. Our observed correlations between precision measures are significantly lower than both the rational simulated benchmarks and the expert-predicted benchmarks for all these correlations. The fact that the average correlation between items in Study 3a's empirical data (average r = .11) remains so much lower than in the (similarly noisy) simulated data (average r = .46) strongly suggests that the difference is not solely attributable to noise.

5.2.2. Correlations between certainty measures and trait measures

Finally, we report correlations between our certainty measures and trait measures in Table 6. We do not find significant correlations between the majority of pairs of certainty measures and trait measures, with the following exceptions: people who scored higher on the Overconfidence Test were more confident on the Likert scale, r = .24, p < .001 and bet more of their bonus, r = .15, p = .024. People who scored higher on actively open-minded thinking had wider confidence interval widths, r = −.18, p = .004, and older people were slightly less confident on the Likert scale, r = −.14, p = .017, dfs = 240.

5.3. Discussion

Correlations between different measures of certainty (average r = .11) are lower than both experts and simulations suggest they should be if they were all tapping the same underlying trait. Instead, we contend that the low correlations between different measures of certainty highlight a fundamental

[3]Participants earned double their bet if they were correct. An agent who bets b should expect to earn $p_i(1+b) + (1-p_i)(1-b) = b(2p_i - 1)+1$. This value increases with b for p_i > .5, and decreases with b for p_i < .5.

[4]Alternatively, we could inject noise by adding a normally distributed variable with standard deviations equal to the empirically observed ones: SD_{Likert} = 1.50, SD_{bet} = 0.35, $SD_{SPD\ peak}$ = 0.18, $SD_{CI\ width}$ = 2.30. This results in similar but slightly higher average simulated correlations (rs .47–.53) than those in Table 5.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 6. *Study 3a: Correlations between certainty measures and trait measures.*

Certainty measure	IH factor 1	IH factor 4	AOT	OCT	Male	Age
SPD peak	.09	.05	.02	.08	.04	.06
Likert	.00	−.04	.05	.24***	.04	−.15*
Bet	−0.04	.01	−.02	.15*	.04	−.02
CI width (reverse-scored)	.04	.00	−.18**	−.04	−.02	−.03

Note: Pearson correlations. IH F1, independence of intellect; IH F4, lack of intellectual overconfidence; AOT, Actively Open-minded Thinking; OCT, Overconfidence Test. Asterisks denote results of two-tailed tests comparing the correlations to 0, *p<.05, **p<.01, ***p<.001.

difference between the simulated and empirical data: the simulated data derive from a coherent underlying understanding of uncertainty. That is, the simulated rational agent holds a subjective probability distribution. Our results suggest that few of our participants build their certainty judgments on such a solid foundation. Instead, people cobble together rough and ready responses to individual responses based on a vague subjective sense of certainty that is not nearly as clear or coherent as a subjective probability distribution. Consequently, small differences in question wording or context can affect responses, making them appear inconsistent. Further, none of the trait measures (actively open-minded thinking, intellectual humility, overconfidence test, demographics) correlate consistently with multiple measures of certainty.

6. Study 3b: Multiple measures of precision

One potential concern about Study 3a is that participants were indifferent about guessing fuzzy images and how many they guessed correctly. If motivation or emotion contributes to confidence and overconfidence, then a trivial task might not be sufficiently ego-involving. Study 3b seeks to address this potential concern. We selected this domain based on the results from the pretest, which identified lay theories that confidence in sports predictions may improve our chances of detecting individual differences in overconfidence; professional sports predictions have also been used in research on related subjects, such as optimism (Simmons and Massey, 2012). In addition, Study 3b uses a more comprehensive suite of trait measures than in Studies 1a–3a, including the Big Five, the need for cognitive closure, and the need for cognition along with measures from previous studies.

6.1. Method

We preregistered this study at https://aspredicted.org/blind.php?x=3DY_792 on May 31, 2023 before data collection started on June 1, 2023.

6.1.1. Participants

We recruited 353 participants from Prolific, using the following filters: participants were located in the US, had an approval rate above 95%, were fluent in English, and had at least 500 previous submissions. 182 participants failed the screening criteria and could not proceed with the survey: participants who answered 'No' to 'Do you identify as an NBA fan?', or who answered any of four multiple-choice questions about the current NBA postseason incorrectly. We excluded two participants who straightlined the actively open-minded thinking scale and six participants whose reported score prediction estimate was more than 20 away from the center of their subjective probability distribution. Our final sample consisted of 163 participants. Our participants were 67.48% male, 68.1% White / 4.29% Asian / 19.02% Black / 1.84% Hispanic / 6.75% Other, M_{age} = 36.53, SD_{age} = 12.28, with amedian education level of a Bachelor's degree in college. Participants earned $2.40 for participating in the study, for which the median completion time was 9.94 min.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

6.1.2. Procedure

On June 1, 2023, participants predicted the winner of the first game of the 2023 NBA Finals (Denver Nuggets or Miami Heat), and by how many points that team would win. All participants completed the survey by 4:30 pm Eastern time, before the game started at 8:30 pm. Participants reported confidence in whether their point prediction was within five points of the true point difference in the same four ways as in Study 3a: a 7-point Likert scale ($M = 4.81$, $SD = 1.28$), an real \$1.00 bet ($M = \$0.53$, $SD = \$0.43$), a 90% confidence interval ($M_{width} = -18.79$, $SD = 20.22$), and a subjective probability distribution ($M_{peak} = 0.55$, $SD_{peak} = 0.19$; $M_{variance} = 347.36$, $SD_{variance} = 354.61$). We also added an additional fifth measure of confidence, a numeric probability between 0 and 100 ($M = 69.56$, $SD = 21.56$) that their prediction was within five points of the true point difference. Participants then reported the following trait measures in a randomized order: actively open-minded thinking, intellectual humility, narcissism, Big Five personality dimensions, need for cognition, and need for cognitive closure. Finally, participants reported their demographics.

6.1.3. Trait measures

Actively open-minded thinking. We used the same scale as in prior studies, $\alpha = 0.79$, $M = 3.84$, $SD = 0.59$.

Intellectual humility, independence of intellect, and ego factor. We used the same scale as in Study 3a, $\alpha = 0.90$, M = 2.53, SD = 0.89.

Intellectual humility, lack of intellectual overconfidence. We used the same scale as in Study 3a, $\alpha = 0.77$, $M = 2.80$, $SD = 0.65$.

Narcissism. We used the same scale as in prior studies, $\alpha = 0.77$, $M = 12.36$, $SD = 3.14$.

Big Five. The BFI-XS measures the Big Five dimensions of personality (Soto and John, 2017). Participants rated the extent to which they agreed with three statements per trait that completed 'I am someone who. . .', 1 = 'disagree strongly' to 5 = 'strongly agree', for extraversion ($\alpha = 0.67$, $M = 2.95$, $SD = 0.93$), openness ($\alpha = 0.66$, $M = 3.91$, $SD = 0.81$), conscientiousness ($\alpha = 0.68$, $M = 3.63$, $SD = 0.93$), agreeableness ($\alpha = 0.59$, $M = 3.89$, $SD = 0.78$), neuroticism ($\alpha = 0.82$, $M = 2.62$, $SD = 1.11$).

Need for Cognition. We used the 6-item Need for Cognition Scale (Lins de Holanda Coelho et al., 2020), for example, 'I would prefer complex to simple problems,' $\alpha = 0.88$, $M = 3.62$, $SD = 0.77$.

Need for Cognitive Closure. We used the 15-item Need for Cognitive Closure scale (Roets and Van Hiel, 2011; Webster and Kruglanski, 1994), for example, 'I would prefer complex to simple problems,' $\alpha = 0.80$, $M = 3.29$, $SD = 0.50$.

6.2. Results

6.2.1. Accuracy

Participants were the most accurate on this task compared to those in previous studies; the majority (123/163, or 75.46%) picked the winning team correctly (the Denver Nuggets). Of those who picked the winning team correctly, the mean prediction for the point difference was 9.68 ($SD = 3.78$), close to the true point difference of 11.

6.2.2. Correlations between certainty measures

Our certainty measures were mostly significantly correlated with each other (see Table 7), with a few exceptions. 90% confidence interval width did not significantly correlate with subjective probability distribution peak, $r = .14$, $p = .073$, or subjective probability distribution variance, $r = -.06$, $p = .458$, nor with bet confidence, $r = -.03$, $p = .729$. In addition, Likert confidence did not correlate with subjective probability distribution variance, $r = -.07$, $p = .367$, though it did correlate with subjective probability distribution peak.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Table 7. *Study 3b: Correlations between certainty measures.*

Variable	SPD peak	SPD Variance	Likert	Bet	90% CI width
SPD peak					
SPD variance	−.74***				
Likert	.21**	−.07			
Bet	.29**	−.29**	.34***		
90% CI width	.14	−.06	.26**	−.03	
Numeric	.26***	−.19**	.34***	.26***	.17*

Note: Asterisks denote results of two-tailed tests comparing the correlations to 0, *p<.05, **p<.01, ***p<.001. Confidence interval width is reverse-scored.

Table 8. *Study 3b: Pearson correlations between certainty measures and trait measures.*

Variable	SPD peak	SPD variance	Likert	Bet	90% CI width	Numeric
Extraversion	.03	.11	.22**	−.01	.11	.05
Openness	−.04	.04	.00	.00	.00	−.02
Neuroticism	−.19*	.06	−.18*	−.04	−.10	−.07
Agreeableness	.08	.00	.10	.05	.06	.16*
Conscientiousness	.16*	−.10	.23**	.00	.24**	.20*
AOT	−.03	−.13	−.20*	.01	−.23**	−.15
IH factor 1	.19*	.18*	−.01	−.03	.00	−.05
IH factor 4	−.05	.20*	.18*	.01	−.06	.04
Narcissism	.00	−.14	−.27***	.06	−.12	−.01
Need for cognition	.00	−.02	.06	.04	−.06	.05
Need for cognitive Closure	−.01	−.01	.12	.04	−.06	.11
Age	.26***	−.14	0.12	.05	.09	.18*
Gender (male)	.12	−.11	.16*	.04	.05	.02

Note: Asterisks denote results of two-tailed tests comparing the correlations to 0, *p<.05, **p<.01, ***p<.001. Confidence interval width is reverse-scored so that higher numbers are narrower intervals.

6.2.3. Correlations between precision measures and trait measures

All correlations between our precision measures and 13 trait measures appear in Table 8. Most correlations between traits and precision measures do not extend to multiple precision measures, with a few exceptions. Actively open-minded thinking negatively correlated with multiple precision measures such that actively open-minded thinkers were generally less confident: with Likert confidence, $r = -.20$, $p = .012$, and confidence interval width (reverse-scored), $r = -.23$, $p = .004$, and directionally negatively correlated with the other precision measures. Interestingly, conscientiousness was positively correlated with multiple precision measures: Likert confidence, $r = .23$, $p = .003$; confidence interval width, $r = .24$, $p = .002$, and numeric probability confidence, $r = .20$, $p = .012$, dfs = 169. We do not see an obvious explanation for why more conscientious people might be more confident, unless they are aware of their own general tendency to be conscientious and successful.

6.3. Discussion

We note that the correlations between certainty measures are generally stronger compared to those in Study 3a; the average of the six correlations between the four precision measures that we also measured in Study 3a (Bet, CI Width, SPD Peak, and Likert) is .20, compared to .13 in Study 3a. This could be due to the nature of the task; participants potentially cared more and were more familiar

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

with sports forecasting than guessing fuzzy images. We find mixed evidence that different measures of overprecision correlate with each other; while single-item continuous scale measures correlated consistently (Likert scale, numeric probability, or a bet), they were less related to the other two elicitation methods that attempt to measure the width or narrowness of one's subjective probability distribution—confidence interval width and probability distribution peak and variance. Study 3b also extends 3a's findings on relationships between certainty measures and trait measures. As in Study 3a, none of the traits we measured correlated consistently with our different measures of precision. However, the one that perhaps seemed most consistent was actively open-minded thinking, which correlated significantly negatively with two precision measures and directionally negatively with all of them. Conscientiousness was positively correlated with four of our six certainty measures; it could be that participants who are willing to answer favorably about their own competence on the conscientiousness measures (e.g., 'I am someone who is reliable, can always be counted on') is also more certain about their competence on this task.

7. General discussion

Overconfidence does not quite exhibit the consistency of a personality trait. In Studies 1a and 1b, we find low within-individual stability for overestimation and overplacement across three domains. In Studies 1b and 2, we find low within-individual stability for overestimation and overplacement across time points on sports forecasting tasks. However, we observe relatively high within-individual stability for overprecision. In Studies 3a and 3b, we find that correlations between different measures of precision produce correlations that are lower than benchmarks predicted by experts and simulations. In sum, we find inconsistency between different forms of overconfidence, between different measures of the same form, between overconfidence in different task domains, and between different points in time. Further, we find weak and inconsistent evidence that trait measures, including both demographic and personality measures, are related to overestimation or overplacement. We cannot attribute these weak correlations solely to poor data quality. Validated scales such as narcissism, the Big Five, and actively open-minded thinking show high reliabilities as expected in our samples, both within and across studies.

Despite these results, we find some hints of consistencies in overprecision. In Study 1a, a single measure of overprecision was correlated (*r*s around 0.7) across multiple domains. In Studies 1b and 2, we find that the correlation between two-time points of overprecision on every task we tested is between 0.4 and 0.5. These results build on past data indicating that overprecision may be the most robust form of overconfidence (Moore, 2023; Moore and Healy, 2008). However, these relationships primarily rely on the measure of overprecision that elicits confidence using a subjective probability distribution; the low correlations between different measures of precision in Studies 3a and 3b raise questions about whether these different measures are actually tapping into the same construct.

We acknowledge that across our studies we do find various significant correlations between confidence, overconfidence, and trait measures. In particular, we find that actively open-minded thinkers seem to be less overconfident in Raven's Progressive Matrices in multiple studies, and are less certain in Study 3b; if there is a trait that correlates with overconfidence or partially explains it, based on our evidence it is likely actively open-minded thinking. Further, we find some evidence that men tend to be more confident across our tasks (though we acknowledge that sports forecasting in particular may be a male-typed task), with less confidence in overconfidence. We find mixed results on other traits. However, we first note that given the sheer number of correlation tests in each study, some are bound to show up as significant. Second, our results are descriptive; while we argue that the magnitude of the relationships between overconfidence and certainty measures across contexts is too low to be considered a trait, readers may interpret these same numerical results differently.

Psychologists seeking to identify new traits must run them through a gauntlet of stringent tests. Their claims gain credibility as their measures demonstrate consistent correlations across diverse samples; over varying periods of time; and using different methods. A measure that withstands these tests

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

earns its place in the pantheon of individual differences following tests of convergent, nomological, and discriminant validity. In our case, these tests are not worth performing. We do not claim that overconfidence deserves a place in that pantheon. Our results show that overconfidence fails the most basic tests of consistency. It makes no sense for us to attempt tests of discriminant validity with other constructs when different measures of overconfidence fail to correlate with one another.

How do we reconcile our results with published claims finding consistent individual differences in overconfidence? We see several possible explanations. First, mono-method bias may inflate the perceived consistency in prior studies; relying on the same single measure of overconfidence might inflate correlations by relying on a single measure, situation or task (Donaldson and Grant-Vallone, 2002). We see mono-method bias as a weakness of our adversarial collaboration on this topic (Binnendyk et al., 2024), which found relatively high correlations between probabilistic item-confidence on various tasks. The present article, by contrast, seeks more stringent measures of generalizability by employing a diversity of methods, approaches, and domains. Second, although we tried to use the best measures of overconfidence that we knew of, the paradigms and measures that we used may be inherently different than those in previous studies. For example, field studies that rely on investment behavior (Bengtsson et al., 2005; Malmendier and Tate, 2005) that lack direct measures of beliefs may tap different internal mechanisms than our measures of overconfidence. Third, we cannot know the extent to which published studies reflect the file-drawering of discrepant results or publication bias; we do know that we have sought to report all the studies that we conducted and err on the side of full disclosure, in this writeup, in our supplemental writeup, and in all our posted documents online.

However, we acknowledge a number of limitations in our data—primarily regarding imperfect measures in Studies 1a, 1b, and 2. Our measures of confidence and overconfidence are single-item measures, which are less reliable than multi-item personality measures; this could deflate correlations with other durable personality traits. In addition, our measures of overplacement and overprecision rely on judgments about other participants in the study, which our participants may have had limited information about. Although we specified in our recruitment materials and consent forms that we wanted MLB (Studies 1a and 1b) or NFL fans (Study 2), we did not explicitly tell participants that all of our participants would be screened and recruited the same way from the same platform. Judgments about other participants were necessary for overplacement, but our measure of overprecision (comparing the spread of participants' estimates of the score distribution to the true spread) was unconventional in this regard. Further, our tasks were relatively difficult; even though we recruited sports 'fans', the participants in Studies 1a and 1b performed no better than chance on the sports forecasting tasks, making the measures of overconfidence more noisy relative to confidence.

Understanding the origins of overconfident judgments is consequential. Evidence suggests that managers who exhibit excessive self-assurance make bad investments (Odean, 1998), issue too much debt (Hackbarth, 2008), and discount useful advice from others (Minson et al., 2011). Overly optimistic entrepreneurs take excessive risks (Hogarth and Karelaia, 2012; Vörös, 2020). Overly optimistic policy makers fail to anticipate and prevent economic crises, like recessions or inflation (Bennani, 2023; Claussen et al., 2012). Overprecision, the most pervasive of the three forms of overconfidence, is evident in nearly every study that compares people's certainty in judgments with the accuracy of those judgments (Campbell and Moore, 2024; Moore, Tenney, et al., 2015). This excessive certainty can serve as the gateway bias, making it harder for people to appreciate the flaws in their judgment and their vulnerability to other biases (Bazerman and Moore, 2012).

Laypeople share the common intuition that there are some people who are more overconfident than others. That assumption undergirds research programs that seek to identify which types of people are more overconfident. Our research tests that assumption and orients researchers to the circumstances in which that assumption does indeed hold. We identify overprecision as the form of overconfidence that is best considered an individual difference, as it is consistent across domains and correlates with other trait measures. This finding opens the way to research identifying how people form their confidence judgments, and therefore, how we can best understand their consistencies and inconsistencies.

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

8. Conclusion

This article evaluates the common belief that overconfidence is an individual difference. We find that overconfidence demonstrates inconsistency across different contexts and time within individuals; further, overestimation and overplacement do not seem to consistently correlate with trait measures. Overprecision, particularly when measured with a subjective probability distribution, shows more consistency across domains, as it correlates with itself across domains and time; however, low consistency between measures of certainty raises questions. We hope to shift the focus of future overconfidence research from identifying overconfident individuals to exploring situational moderators that influence confidence judgments.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/jdm.2025.11.

Data availability statement. Raw and cleaned data for all studies are available at https://osf.io/tb2me/.

Acknowledgments. This article benefited from helpful comments by Rene Choudhari, Karin Garrett, Aryan Arora, and Angelica Wang.

Funding statement. This research received no specific grant funding from any funding agency, commercial or not-for-profit sectors. The authors gratefully acknowledge funding from the Dean's Office at Berkeley-Haas and the Experimental Social Science Laboratory at UC Berkeley.

Competing interest. The authors declare no competing interests.

References

Acker, D., & Duck, N. W. (2008). Cross-cultural overconfidence and biased self-attribution. *Journal of Socio-Economics*, *37*(5), 1815–1824.

Ackerman, R., & Thompson, V. A. (2017). meta-reasoning: monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Ames, D. R., & Kammrath, L. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior*, *28*, 187–209.

Ames, D. R., Rose, P., & Anderson, C. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, *40*(4), 440–450.

Barber, B. M., & Odean, T. (2001). boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, *116*(1), 261–292. https://doi.org/10.1162/003355301556400

Bazerman, M. H., & Moore, D. A. (2012). *Judgment in managerial decision making*. John Wiley & Sons.

Bengtsson, C., Persson, M., & Willenhag, P. (2005). Gender and overconfidence. *Economics Letters*, *86*(2), 199–203. https://doi.org/10.1016/j.econlet.2004.07.012

Bennani, H. (2023). Overconfidence of the chair of the federal reserve and market expectations: Evidence based on media coverage. *International Journal of Finance & Economics*, *28*(3), 3403–3419. https://doi.org/10.1002/ijfe.2599

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, *59*(5), 960–970.

Binnendyk, J., & Pennycook, G. (2024). Individual differences in overconfidence: A new measurement approach. *Judgment and Decision Making*, *19*, e28. https://doi.org/10.1017/jdm.2024.22

Binnendyk, J., Li, S., Costello, T., Hale, R., Moore, D. A., & Pennycook, G. (2024). Is overconfidence a trait? An adversarial collaboration. PsyArXiv. https://doi.org/10.31234/osf.io/awugz

Bornstein, B. H. & Zickafoose, D. J. (1999). I know I know it, I know I saw it: The stability of the confidence–accuracy relationship across domains. *Journal of experimental psychology: Applied*, *5*(1), 76.

Bors, D. A. & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*(3), 382–398.

Bowes, S. M., Ringwood, A., & Tasimi, A. (2024). Is intellectual humility related to more accuracy and less overconfidence? *The Journal of Positive Psychology*, *19*(3), 538–553.

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, *67*(1), 59–77. https://doi.org/10.1016/j.jml.2012.04.002

Campbell, S., & Moore, D. A. (2024). Overprecision in the survey of professional forecasters. *Collabra: Psychology*, *10*(1), 92953. https://doi.org/10.1525/collabra.92953

Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making*, *17*(4), 297–311. https://doi.org/10.1002/bdm.475

Claussen, C. A., Matsen, E., Røisland, Ø., & Torvik, R. (2012). Overconfidence, monetary policy committees and chairman dominance. *Journal of Economic Behavior & Organization*, *81*(2), 699–711.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgements: A calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences*, *8*(2), 83–103.

Dahlbom, L., Jakobsson, A., Jakobsson, N., & Kotsadam, A. (2011). Gender and overconfidence: Are girls really overconfident? *Applied Economics Letters*, *18*(4), 325–327. https://doi.org/10.1080/13504851003670668

Deaves, R., Lüders, E., & Schröder, M. (2010). The dynamics of overconfidence: Evidence from stock market forecasters. *Journal of Economic Behavior & Organization*, *75*(3), 402–412. https://doi.org/10.1016/j.jebo.2010.05.001

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245–260. https://doi.org/10.1023/A:1019637632584

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences* (pp. xxii, 600). Thomson Brooks/Cole Publishing Co.

Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381. https://doi.org/10.1093/qje/qjac003

Fleeson, W., & Gallagher, P. (2009). The implications of big five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168.

Gardner, H., & Hatch, T. (1989). Educational implications of the theory of multiple intelligences. *Educational Researcher*, *18*(8), 4–10. https://doi.org/10.3102/0013189X018008004

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Glaser, M., Langer, T., & Weber, M. (2005). Overconfidence of professionals and lay men: Individual differences within and between tasks? *Unpublished Manuscript*. https://core.ac.uk/download/pdf/6321722.pdf

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*(1), 1–14. https://doi.org/10.1037/e513702014-109

Hackbarth, D. (2008). Managerial traits and capital structure decisions. *Journal of Financial and Quantitative Analysis*, *43*(4), 843–881.

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L.-G. (2008). Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and Aging*, *23*(3), 531–544.

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*(7), 467–476. https://doi.org/10.1037/e615882011-200

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*(3), 188–201. https://doi.org/10.1017/S1930297500005921

Hogarth, R. M., & Karelaia, N. (2012). Entrepreneurial success and failure: Confidence and fallible judgment. *Organization Science*, *23*(6), 1733–1747.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters*, *110*(2), 97–100. https://doi.org/10.1016/j.econlet.2010.11.015

Jackson, S. A., Kleitman, S., Howie, P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01559

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). Continuous univariate distributions, volume 2 (Vol. 2). John wiley & sons.

Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, *34*(4), 559–574. https://doi.org/10.1016/S0191-8869(02)00028-4

Kahneman, D. (2011). *Thinking fast and slow*. Farrar, Straus and Giroux.

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216–247. https://doi.org/10.1006/obhd.1999.2847

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118.

Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. *Journal of Personality Assessment*, *98*(2), 209–221.

Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, *76*(2), 315–332.

Langnickel, F., & Zeisberger, S. (2016). Do we measure overconfidence? A closer look at the interval production task. *Journal of Economic Behavior & Organization*, *128*, 121–133.

Larkin, I., & Leider, S. (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics*, *4*(2), 184–214.

Lawson, M. A., Larrick, R. P., & Soll, J. B. (2023). Forms of overconfidence: Reconciling divergent levels with consistent individual differences. *SSRN Scholarly Paper 4558486*. https://doi.org/10.2139/ssrn.4558486

Lins de Holanda Coelho, G., HP Hanel, P., & J Wolf, L. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Littrell, S., Risko, E. F., & Fugelsang, J. A. (2021). 'You can't bullshit a bullshitter' (or can you?): Bullshitting frequency predicts receptivity to various types of misleading information. *British Journal of Social Psychology*, *60*(4), 1484–1505. https://doi.org/10.1111/bjso.12447

Logg, J. M., Haran, U. & Moore, D. A. (2018). Is overconfidence a motivated bias? Experimental evidence. *Journal of Experimental Psychology: General*, *147*(10), 1445.

Lundeberg, M. A., Fox, P. W., & Punćcohaŕ, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*, 114–121. https://doi.org/10.1037/0022-0663.86.1.114

Malmendier, U., & Tate, G. (2005). Does overconfidence affect corporate investment? CEO overconfidence measures revisited. *European Financial Management*, *11*(5), 649–659.

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, *22*(2), 274–281.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*(1), 1–14. https://doi.org/10.1037/xap0000040

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325–1338.

Mischel, W. (1968). *Personality and assessment*. Wiley.

Moore, D. A. (2023). Overprecision is a property of thinking systems. *Psychological Review*, *130*(5), 1339–1350. https://doi.org/10.1037/rev0000370

Moore, D. A., Carter, A., & Yang, H. H. J. (2015). Wide of the mark: Evidence on the underlying causes of overprecision in judgment. *Organizational Behavior and Human Decision Processes*, *131*, 110–120.

Moore, D. A., & Dev, A. S. (2017). Overconfidence. In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_1157-1

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

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, *11*(8). https://doi.org/10.1111/spc3.12331

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, *92*(6), 972–989. https://doi.org/10.1037/0022-3514.92.6.972

Moore, D. A., & Swift, S. A. (2010). The three faces of overconfidence in organizations. In R. Van Dick & J. K. Murnighan (Eds.), *Social psychology of organizations* (pp. 147–184). Taylor & Francis.

Moore, D. A., Tenney, E. R., & Haran, U. (2015). Overprecision in judgment. In G. Wu & G. Keren (Eds.), *Handbook of judgment and decision making* (pp. 182–212). Wiley.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*(3), 1067–1101.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, *53*(6), 1887–1934.

O'Reilly, C. A., & Hall, N. (2021). Grandiose narcissists and decision making: Impulsive, overconfident, and skeptical of experts–but seldom in doubt. *Personality and Individual Differences*, *168*, 110280. https://doi.org/10.1016/j.paid.2020.110280

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, *129*(3), 257. https://doi.org/10.1080/00221300209602099

Pennycook, G., Binnendyk, J., & Rand, D. G. (2022). Overconfidently conspiratorial: Conspiracy believers are dispositionally overconfident and massively overestimate how much others agree with them [Preprint]. *PsyArXiv*. https://osf.io/d5fz2

Prims, J. P., & Moore, D. A. (2017). Overconfidence over the lifespan. *Judgment and Decision Making*, *12*(1), 29–41.

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and individual differences*, *50*(1), 90–94.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the big five. *Journal of Research in Personality*, *38*(5), 473–480.

Shariatmadari, D. (2015). Daniel Kahneman: 'What would I eliminate if I had a magic wand? Overconfidence.' *The Guardian*. https://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview

Simmons, J. P., & Massey, C. (2012). Is optimism real?. *Journal of Experimental Psychology: General*, *141*(4), 630.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81.

Spearman, C. (1961). *"General Intelligence" Objectively determined and measured* (p. 73). Appleton-Century-Crofts. https://doi.org/10.1037/11491-006

Spiller, S. A. (2024). Widely-used measures of overconfidence are confounded with ability. *SSRN Scholarly Paper 4468920*. https://doi.org/10.2139/ssrn.4468920

https://doi.org/10.1017/jdm.2025.11 Published online by Cambridge University Press

Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, *21*(6), 971–986. https://doi.org/10.1016/S0191-8869(96)00130-4

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, *22*(4), 259–264. https://doi.org/10.1177/0963721413480174

Sternberg, R. J. (2021). Adaptive intelligence: Intelligence is not a personal trait but rather a person × Task × Situation Interaction. *Journal of Intelligence*, *9*(4), 4. https://doi.org/10.3390/jintelligence9040058

Vörös, Z. (2020). Effect of the different forms of overconfidence on venture creation: Overestimation, overplacement and overprecision. *Journal of Management and Organization*. https://www.cambridge.org/core/journals/journal-of-management-and-organization/article/effect-of-the-different-forms-of-overconfidence-on-venture-creation-overestimation-overplacement-and-overprecision/73D0C2B067FB6F4E1EA7134E62C6A7C1

Walters, D. J., Fernbach, P. M., Fox, C. R., & Sloman, S. A. (2017). Known unknowns: A critical determinant of confidence and calibration. *Management Science*, *63*(12), 4298–4307. https://doi.org/10.1287/mnsc.2016.2580

Weber, E. U., & Johnson, E. J. (2009). Decisions under uncertainty: Psychological, economic, and neuroeconomic explanations of risk preference. In *Neuroeconomics* (pp. 127–144). Elsevier.

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and Lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430–445. https://doi.org/10.1037/0033-295X.111.2.430

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in the need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049–1062.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review*, *4*(3), 387–392. https://doi.org/10.3758/BF03210798

Cite this article: Li, S., Hale, R., and Moore, D. A. (2025). Is overconfidence an individual difference? *Judgment and Decision Making*, e25. https://doi.org/10.1017/jdm.2025.11