THE JOURNAL OF POSITIVE PSYCHOLOGY
2024, VOL. 19, NO. 3, 538–553
https://doi.org/10.1080/17439760.2023.2208100

Routledge
Taylor & Francis Group

Is intellectual humility related to more accuracy and less overconfidence?

Shauna M. Bowes, Anna Ringwood and Arber Tasimi

Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA, USA

ABSTRACT

Is intellectual humility (IH) related to more accuracy and less overconfidence in decision-making contexts? Here we sought to answer this question and clarify ambiguities in the literature by examining the relations among IH, critical thinking, and overconfidence measures. We assessed these relations in both online community and college participants. Overall, IH tended to be related to more accuracy on a range of critical thinking measures and to less overestimation and overclaiming. We also found some evidence for a Dunning-Kruger effect (i.e. those scoring the lowest on IH significantly overestimated their performance). Results tended to be significantly stronger in the online community participants than in the college participants, and the Dunning-Kruger effect was not present in the college sample. Overall, the current findings invite questions about how IH contributes to more accuracy and less overconfidence and whether IH can be leveraged to bolster critical thinking.

ARTICLE HISTORY
Received 2 September 2022
Accepted 16 March 2023

KEYWORDS
Intellectual humility; critical thinking; overconfidence; Dunning-Kruger effect; humility

Overconfidence – the propensity to exhibit more confidence than is justified – can lead to disastrous consequences, from medical errors (e.g. Berner & Graber, 2008) to major historical calamities (e.g. the Chernobyl nuclear accident; see Moore & Schatz, 2017). Not only is overconfidence potentially dangerous, but it is also pervasive, with most people failing to recognize signs of overconfidence in themselves (e.g. Anderson et al., 2012).

Given that it is both consequential and widespread, scholars have attempted to identify ways to mitigate overconfidence for decades (e.g. Larrick, 2004; Sellier & Scopelliti, 2019). Most strategies for reducing overconfidence can be subsumed by the umbrella category of *debiasing strategies*, which collectively aim to increase people's awareness of bias and decrease reliance on faulty heuristics (e.g. Lewandowsky et al., 2012). Debiasing strategies tend to include education on bias (e.g. to increase awareness of bias; see Jenkins & Youngstrom, 2016) and/or strategies for countering bias once it is recognized (e.g. using checklists to decrease reliance on memory; see Arkes, 1991). In essence, debiasing strategies involve building awareness of the existence of bias in one's decision-making and one's abilities to make alternative choices.

Importantly, the first step for any debiasing strategy to be effective is for individuals to possess the motivation and desire to self-reflect. That is, to be less overconfident, one must recognize that there is a possibility that they may be wrong. Thus, understanding whether *intellectual humility* (IH)—the propensity to self-reflect and reconsider one's views when evidence points in other directions (e.g. Porter & Schumann, 2018)––is related to more accuracy and less overconfidence is an open and interesting question.

To our knowledge, only one study has directly examined the links between IH and both accuracy and overconfidence (Leman et al., 2021). This work examined whether IH moderated the Dunning-Kruger effect (DK effect), a psychological phenomenon indicating that people with the lowest level of ability on a particular task (e.g. algebra problems) or in a particular domain (e.g. mathematics) are the most overconfident about their ability (e.g. Kruger & Dunning, 1999). This study investigated whether those scoring at the lowest quartiles of IH were the most likely to overestimate their performance on measures of intelligence and general knowledge (e.g. health, geography). Overall, there was little evidence that IH was related to objective performance on critical thinking measures (accuracy) and for a DK effect (overconfidence).

At first blush, it may seem like IH is related to neither accuracy nor overconfidence based on the above findings, yet there are several pieces of evidence that speak against this possibility. For one, IH tends to be significantly and positively related to multiple measures of accuracy, including intelligence (e.g. Bowes et al., 2022; Zmigrod et al., 2019), cognitive

CONTACT Shauna M. Bowes shauna.m.bowes@gmail.com Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA 30322, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/17439760.2023.2208100.

flexibility (e.g. Zmigrod et al., 2019), cognitive reflection (e.g. Krumrei-Mancuso et al., 2020), science literacy (e.g. Bowes et al., 2022), metamemory (Deffler et al., 2016), and the ability to distinguish between true and false information (e.g. Bowes & Tasimi, 2022). With respect to overconfidence, some research indicates that intellectually humble individuals are less likely to claim they are familiar with topics that do not exist (Alfano et al., 2017; Krumrei-Mancuso et al., 2020).

Here we sought to address these discrepancies by taking another look at the relations between IH and accuracy and overconfidence. Although our approach was largely inspired by previous work (e.g. Leman et al., 2021), the current investigation is broader in several ways. First, we employed different measures of IH, as measures of IH vary in their coverage of potential features of IH based on their differing conceptualizations of the construct. For instance, all measures of IH assess metacognitive aspects of IH, such as the tendency to self-reflect and remain open-minded to alternatives. Indeed, some measures, like the *General Intellectual Humility Scale* (Leary et al., 2017), only assess metacognitive aspects of IH. Other measures, like the *Comprehensive Intellectual Humility Scale* (Krumrei-Mancuso & Rouse, 2016), adopt a broader view of IH, including metacognitive items as well as items assessing emotional (e.g. distress tolerance) and interpersonal (e.g. respect for others' views) aspects of IH. These differences in IH conceptualizations and measures can translate into differences in their relations with external criteria, including perhaps accuracy and overconfidence. For example, research has found that measures assessing a blend of metacognitive, emotional, and interpersonal IH features are stronger correlates of less susceptibility to misinformation and conspiratorial ideation than measures assessing metacognitive features alone (Bowes & Tasimi, 2022). Thus, it is possible that the relations between IH and overconfidence may vary across measures of IH that differ in their coverage of the metacognitive, interpersonal, and emotional aspects of IH.

Similarly, some research has found that domain-specific measures of IH (i.e. IH for a specific belief, topic, or issue; see Hoyle et al., 2016) are stronger correlates of relevant external criteria than domain-general (i.e. decontextualized) measures of IH. For instance, politics-specific IH is a stronger correlate of less political myside bias than domain-general IH (e.g. Bowes et al., 2022). It is possible then that domain-specific IH, which has not yet been examined in relation to accuracy and overconfidence, may be a stronger correlate of certain forms of accuracy and overconfidence than domain-general IH. By including these measures, it will be possible to identify boundaries in the relations among IH, accuracy, and overconfidence.

Finally, we examined whether any potential relations between IH and both accuracy and overconfidence were unique to IH or to humility more generally. Recent research indicates that modesty, a facet of humility, is related to more accurate self-assessment on decision-making tasks (e.g. Teo et al., 2022). Perhaps IH is related to more accuracy and less overconfidence due to its shared variance with humility at large. By measuring IH alongside general humility, it will be possible to clarify the extent of overlap between the two constructs in the context of accuracy and overconfidence.

Taken together, the current study addressed the following four questions: (1) Is IH related to more accuracy across IH and critical thinking measures? (2) Is IH related to less overconfidence across IH and critical thinking measures? (3) Are relations consistent across IH measures and conceptualizations? (4) Is IH uniquely related to more accuracy and less overconfidence compared with humility at large?

Methods

Our research plan and analyses were not pre-registered. This study was approved by the Emory University Institutional Review Board, and all participants provided informed consent.

Participants

We recruited more than 400 participants in each of four samples to have at least 80% power to detect a medium effect size (i.e. $r = .20$; Gignac & Szodorai, 2016). We screened for inattentive and careless responding using several methods in each sample (see Supplemental Materials 1). We included both online community and college participants.

Online community sample

We collapsed across 3 samples that were recruited from Amazon's Mechanical Turk (Samples 1–3; $N = 1{,}502$). Most participants in the sample (age range = 19–81 years; $M_{age} = 39.10$, $SD_{age} = 11.83$) identified as female (55.5%) and reported having a Bachelor's degree (38.6%). In addition, most participants identified as White (80.6%), followed by Black or African-American (12.6%) and Hispanic (8.7%). The characteristics of each sample can be found in Supplemental Materials 2. The measures we employed varied across samples (see Table 1), and the sample sizes for our analyses ranged from 498 to 1,502.

Table 1. Overview of measures across samples.

	Online community sample			College sample
	Sample 1 ($N = 527$)	Sample 2 ($N = 498$)	Sample 3 ($N = 477$)	Sample 4 ($N = 596$)
Intellectual Humility				
General IH Scale (GIHS)	✓	✓	✓	✓
Porter IH Scale (PIHS)	✓	-	-	✓
Comprehensive IH Scale (CIHS)	✓	✓	✓	✓
McElroy IH Scale (MIHS)	✓	-	-	✓
Alfano IH Scale (AIHS)	✓	-	-	✓
Politics-Specific IH Scale (SIHS)	-	✓	✓	-
Critical thinking Measures				
Intelligence (ICAR)	✓	✓	✓	✓
Trivia Knowledge	✓	-	-	✓
Political Knowledge	-	✓	✓	-
Cognitive Reflection (CRT)	✓	✓	✓	✓
Bullshit Receptivity (BSR)	✓	-	-	✓
Overclaiming	✓	✓	-	✓
Honesty-Humility	✓	✓	✓	✓

College sample

College students enrolled in an introductory psychology course at a private university in the southeast were also recruited ($N = 592$; Sample 4). The sample (age range = 18–26 years; $M_{age} = 18.90$, $SD_{age} = 1.05$) was primarily female (67.8%). Most participants identified as White (51.5%), followed by Asian (34.5%) and Hispanic (11.4%).

Measures

Participants completed an online battery of self-report measures and critical thinking tasks. Other individual differences measures (e.g. cognitive styles, partisanship) were included in this dataset, but they were not analyzed as a part of this study and are being used in ongoing research. Datasets and relevant measures/materials are available at https://osf.io/w4fhv/?view_only=d77f83342e1c410eb5f5f8e39c4a1858v.

Descriptive statistics, internal consistency coefficients, and item response scales are reported in Table 2. Intercorrelations amongst measures are reported in Supplemental Tables S1-S3.

Intellectual humility

In Samples 1 and 4, participants completed five self-report measures of IH. Two of the five measures were unidimensional, yielding total scores: (a) the *General Intellectual Humility Scale* (GIHS; Leary et al., 2017), a 6-item measure of the metacognitive components of IH, and (b) the *Porter Intellectual Humility Scale* (PIHS; Porter & Schumann, 2018), a 9-item measure of the metacognitive, interpersonal, and emotional components of IH.

The remaining three measures in Samples 1 and 4 were multidimensional, assessing the metacognitive, interpersonal, and emotional components of IH: (a) the *Comprehensive Intellectual Humility Scale* (CIHS; Krumrei-Mancuso & Rouse, 2016), a 22-item measure of IH, (b) the *McElroy Intellectual Humility Scale* (MIHS; McElroy et al., 2014), a 16-item measure of IH, and (c) the *Alfano Intellectual Humility Scale* (AIHS; Alfano et al., 2017), a 23-item measure of IH.[1] The subdimensions from these measures are described in Supplemental Materials 4; herein, we focus on the total scores in our analyses.

In Samples 2 and 3, participants completed three self-report measures of IH: the CIHS, the GIHS, and the *Specific Intellectual Humility Scale* (Hoyle et al., 2016) in the domain of politics, a 9-item measure of politics-specific IH (SIHS). The SIHS is unidimensional.

For analyses where we collapse across domain-general IH total scores, refer to Supplemental Table S8. We included these various measures of IH because, as previously mentioned, there is no agreed upon conceptualization of IH.

Critical thinking

We assessed critical thinking using performance-based and knowledge-based measures, including measures of intelligence, trivia knowledge, political knowledge, cognitive reflection, and bullshit receptivity.

Intelligence. Participants in all samples completed a 16-item version of *The International Cognitive Ability Resource* (ICAR; Condon & Revelle, 2014), which is a public-domain measure of intelligence. We focus on the ICAR total score in the main analyses, but results including the ICAR subdimensions are reported in Supplemental Table S4. Participants had 15 minutes to complete the ICAR.

Trivia knowledge. Participants in Samples 1 and 4 completed a 15-item measure of trivia knowledge that spanned

Table 2. Internal consistency statistics and descriptive statistics for study constructs.

	Mean (SD)		α		
	Online Community	College	Online Community	College	Item response scale
GIHS	23.45 (4.20)	24.01 (3.47)	.87	.89	1 (*strongly disagree*) to 5 (*strongly agree*)
CIHS	82.29 (12.83)	83.25 (9.91)	.90	.86	1 (*strongly disagree*) to 5 (*strongly agree*)
PIHS	46.70 (7.93)	47.39 (6.65)	.75	.72	1 (*strongly disagree*) to 7 (*strongly agree*)
MIHS	60.43 (9.85)	60.80 (8.01)	.88	.86	1 (*strongly disagree*) to 5 (*strongly agree*)
AIHS	106.68 (19.19)	104.42 (15.34)	.89	.87	1 (*strongly disagree*) to 7 (*strongly agree*)
SIHS	31.56 (7.87)	-	.91	-	1 (*not at all like me*) to 5 (*very much like me*)
Intelligence Total	7.08 (3.57)	8.88 (2.95)	-[1]	-[1]	Correct responses = 1, incorrect responses = 0
Intelligence Confidence	50.33 (26.98)	49.28 (25.31)	-	-	0% (*not at all confident*) to 100% (*extremely confident*)
Intelligence Perceived Correct	7.62 (3.97)	8.10 (3.57)	-	-	-
Intelligence Self-placement	48.60 (24.05)	51.18 (21.86)	-	-	1st to 99th percentile
Intelligence Overestimation	.51 (3.72)	-.77 (3.05)	-	-	-
Intelligence Overplacement	-1.44 (24.08)	1.18 (21.86)	-	-	-
Trivia Knowledge Total	7.46 (2.81)	6.60 (2.10)	-[3]	-[3]	Correct responses = 1, incorrect responses = 0
Trivia Confidence	48.23 (27.56)	33.35 (20.55)	-	-	0% (*not at all confident*) to 100% (*extremely confident*)
Trivia Perceived Correct	6.89 (3.44)	5.12 (2.65)	-	-	-
Trivia Self-Placement	49.20 (22.61)	44.15 (19.68)	-	-	1st to 99th percentile
Trivia Overestimation	-.57 (3.10)	-1.50 (2.75)	-	-	-
Trivia Overplacement	-.84 (22.60)	-5.85 (19.68)	-	-	-
Political Knowledge Total	3.13 (1.28)	-	-[3]	-[3]	Correct responses = 1, incorrect responses = 0
Political Knowledge Confidence	64.30 (29.11)	-	-	-	0% (*not at all confident*) to 100% (*extremely confident*)
Political Knowledge Perceived Correct	3.42 (1.41)	-	-	-	-
Political Knowledge Self-Placement	57.74 (24.84)	-	-	-	1st to 99th percentile
Political Knowledge Overestimation	.26 (1.35)	-	-	-	-
Political Knowledge Overplacement	-50.01 (.95)	-	-	-	-
Cognitive Reflection	2.11 (1.26)	1.14 (1.06)	-	-	Correct responses = 1, incorrect responses = 0
Bullshit receptivity *d'*	-[2]	-[2]	-[2]	-[2]	1 (*not at all profound*) to 5 (*very profound*)
Bullshit receptivity *c*	-[2]	-[2]	-[2]	-[2]	-
Overclaiming *d'*	-[2]	-[2]	-[2]	-[2]	0 (*never heard of it*) to 6 (*very familiar*)
Overclaiming *c*	-[2]	-[2]	-[2]	-[2]	-
Honesty-Humility	-[4]	52.41 (9.85)	.86	.81	1 (*strongly disagree*) to 5 (*strongly agree*)

Note. GIHS = General Intellectual Humility Scale; CIHS = Comprehensive Intellectual Humility Scale; PIHS = Porter Intellectual Humility Scale; MIHS = McElroy Intellectual Humility Scale; AIHS = Alfano Intellectual Humility Scale; SIHS = Politics-Specific Intellectual Humility Scale.

[a]The Cronbach's alpha coefficient for the ICAR was not calculated, as it is not an appropriate metric for reliability with timed administrations (Cronbach & Shavelson, 2004).

[b]These measures were analyzed using Signal Detection Theory analyses; hence, we did not calculate composites for these measures.

[c]Based on recommendations in the literature, we did not calculate Cronbach's alpha coefficients for the knowledge and performance-based measures (Taber, 2018).

[d]The mean is 0 and the standard deviation is 1, as the Honesty-Humility composite was standardized to account for the fact that a 60-item version and 100-item version of the HEXACO were used.

multiple domains (e.g. geography, entertainment). All questions were multiple-choice with 4 response options.

Political knowledge. Participants in Samples 2 and 3 completed a 5-item measure of political knowledge (Anson, 2018). All questions were multiple-choice with 2 to 5 response options.

Cognitive reflection. Participants in all samples completed the *Cognitive Reflection Test* (CRT; Frederick, 2005), which assesses one's ability to override an incorrect intuitive response and engage in analytical reflection to obtain the correct response. In Samples 1 and 4, we used a 3-item version (Frederick, 2005), and in Samples 2 and 3, we used a 4-item version (Thompson & Oppenheimer, 2016).

Bullshit receptivity. Participants in Samples 1 and 4 completed the *Profound Statements Questionnaire* (PSQ; Pennycook et al. 2015), a self-report measure of one's ability to discriminate semantically meaningless sentences (e.g. 'Wholeness quiets infinite phenomena') from semantically meaningful sentences (e.g. 'Only those who will risk going too far can possibly find out how far one can go'). There were 10 meaningless and 10 meaningful sentences.

We analyzed the PSQ using Signal Detection Theory (SDT). We calculated *d'* (accuracy) and *c* (response bias; see Paulhus et al., 2003); additional details about scoring for the PSQ are in Supplemental Materials 3. Regarding *d'*, the larger the accuracy parameter, the greater the differentiation between profound and meaningless items. Regarding *c*, negative values reflect the tendency

to endorse an item as profound, regardless of whether the item is meaningful or meaningless.

Overconfidence

We assessed overconfidence concerning one's performance on the ICAR, trivia knowledge questionnaire, and political knowledge questionnaire.

Overestimation. Participants indicated how many questions they believe they answered correctly out of the total number of items on the questionnaire. To generate an overestimation score on each measure, we subtracted the actual number of correct items from the perceived number of correct items (i.e. positive scores indicate overestimation whereas negative scores indicate underestimation).

Overplacement. Participants were shown an image of a normal distribution and asked to indicate how well they thought performed, in terms of percentiles, relative to the average person scoring at the 50th percentile (1st to 99th percentile). To generate an overplacement score, we subtracted 50 (representing the 50th percentile) from the participant's self-placement rating (i.e. positive scores indicate overplacement whereas negative scores indicate under-placement).

Confidence. Participants also indicated how confident they were in the accuracy of their answers overall on a 0% (*not at all confident*) to 100% (*extremely confident*) sliding scale.

Overclaiming

Participants in Samples 1, 3, and 4 completed a 45-item version of the *Overclaiming Questionnaire* (OCQ; Paulhus et al., 2003), which comprised 33 'real' items (e.g. photon) and 12 'foil' items (e.g. ultra-lipid).[2] The foil items were generated such that they seem real but in fact do not exist. We analyzed the OCQ using SDT (Paulhus et al., 2003); additional details about scoring for the OCQ are reported in Supplemental Materials 3. Regarding *d'*, the larger the accuracy parameter, the greater the differentiation between real and foil items. Regarding *c*, negative values reflect the tendency to endorse an item as familiar, regardless of whether the item is real or a foil.

Honesty-humility

Participants in Samples 1, 2, and 4 completed the 100-item version of HEXACO PI-R (Lee & Ashton, 2018); in Sample 3, participants completed the 60-item version of the HEXACO PI-R (Ashton & Lee, 2009). Our analyses focus on the Honesty-Humility dimension.

Data analytic plan

We examined the zero-order correlations between IH and critical thinking and overconfidence. All effect sizes were interpreted according to Gignac and Szodorai's (2016) guidelines for individual differences research. We employed Hommel's correction method within IH measure and within sample to account for multiple comparisons.

Of 65 correlations between IH measures and indices of critical thinking and overconfidence, sample population (online community sample vs. college sample) statistically moderated 21 (32%) of these relations. All of the significant moderation results suggested that the relations between IH and critical thinking and overconfidence were stronger in the online community sample than in the college sample. That is, the (a) positive correlations between IH and critical-thinking were stronger and (b) negative correlations between IH and overconfidence were stronger in the online community sample than in the college sample. As such, we separated the online community sample from the college sample.

To compare the relations between IH and critical thinking and overconfidence between the two samples, we conducted tests of the difference between independent correlations using Fisher's *r*-to-*z* transformation and interpreted significance based on two-tailed *p*-values (Preacher, 2002).[3] In addition, we examined the potential for a Dunning-Kruger effect in the relations between IH and critical thinking. We generated plots using the Loess method of participants' perceived accuracy and their objective accuracy on each problem-solving measure (e.g. ICAR) at quartiles of IH. To facilitate interpretation, all missing cases on measures included in the figures (e.g. CIHS and ICAR) were excluded prior to generating the figures. To test whether there were significant differences between objective and perceived scores at each IH percentile, we conducted paired (dependent) samples *t*-tests and calculated Cohen's *d* statistics.

Results

The results for each individual sample are reported in Supplemental Table S5 and Supplemental Figures S1-S3. The correlations between IH subdimensions and critical thinking and overconfidence are reported in Supplemental Tables S6–S7; the Dunning-Kruger plots for each IH subdimension are available at https://osf.io/w4fhv/?view_only=d77f83342e1c410eb5f5f8e39c4a1858v.

Correlations between IH and critical thinking

Online community sample

IH tended to manifest small-to-medium positive correlations with critical thinking measures (Table 3), meaning that IH was generally correlated with more accuracy. In general, IH measures that assessed primarily metacognitive features (e.g. the GIHS) were largely similar in their relations with accuracy to IH measures that assessed a blend of metacognitive, interpersonal, and emotional features (e.g. the CIHS).

IH tended to be weakly-to-moderately related to more intelligence (*rs* ranged from .06 [PIHS] to .22 [CIHS]), weakly related to more trivia knowledge (*rs* ranged from .05 [MIHS] to .13 [GIHS & PIHS]), and weakly related to more cognitive reflection (*rs* ranged from .07 [PIHS & SIHS] to .17 [CIHS]). Regarding political knowledge, the GIHS and CIHS were weakly related to more political knowledge, although the relationship was only significant for the CIHS (*rs* were .08 and .13); in contrast, politics-specific IH was weakly related to less political knowledge, although this relationship was not significant ($r = -.05$). IH was additionally weakly-to-moderately associated with more accuracy (*d'*) on the bullshit receptivity measure, meaning that IH was related to more of an ability to discriminate between meaningless and meaningful statements (*rs* ranged from .12 [PIHS] to .24 [CIHS]). There was little evidence to suggest that IH was linked with response bias, meaning that IH was not systematically related to tendencies to rate items as more or less profound across items (*rs* ranged from -.04 [PIHS] to .16 [AIHS]).

College sample

IH was generally not significantly related to critical thinking (and, thus, accuracy), although most correlations were still positive and small (Table 3). Even though there were few differences across IH measures in their relations with critical thinking, the exception was the PIHS. The PIHS tended to yield the smallest relations with critical thinking. Moreover, it was the only domain-general measure of IH to be related to less critical thinking, specifically on the intelligence and cognitive reflection measures.

No IH measures were significantly related to intelligence or cognitive reflection, and only the AIHS was significantly correlated with trivia knowledge. There was more evidence for significant relations on the bullshit receptivity measure. Specifically, IH measures tended to be weakly and significantly associated with more accuracy (*d'*) on the bullshit receptivity measures (*rs* ranged from -.03 [PIHS] to .17 [AIHS]). There was again little evidence to suggest that IH was linked with response bias on the bullshit receptivity measure.

Sample differences

There were significant differences between samples, as 8 of 25 correlations were significantly different (32% of the results, $ps < .05$) between the online community and college samples. All of the significant results indicated

Table 3. Correlations between IH total scores and critical thinking measures and overconfidence.

	GIHS		CIHS		PIHS		MIHS		AIHS		SIHS		Honesty-Humility	
	Com.	College	Com.	College	Com.	College	Com.	College	Com.	College	Com.	College	Com.	College
Intelligence	**.19**[b]	.11	**.22**[a,b]	.11	.06[a]	-.08[b]	**.15**[a]	.02	**.17**	.11[b]	**.12**	-	**.09**	.06
Confidence	.01	.10[b]	**-.09**[a]	.07	-.07	-.08	-.08[a]	.06	-.10[a]	.07	-.01	-	-.05	-.02
Overestimation	**-.11**[a]	.07	**-.24**[a,b]	.04	-.04	.01	-.12	.07[b]	**-.16**	-.01	-.08	-	**-.11**	-.03
Overplacement	.03[b]	.09[b]	-**.09**[a]	.10[b]	-.10[a]	-.06	-.11[a]	.06[b]	-.09[a]	.03	-.01	-	**-.08**	-.04
Pol. Know.	.08[b]	-	**.13**	-	-	-	-	-	-	-	-.05[b]	-	**.15**	-
Confidence	.02	-	-.04	-	-	-	-	-	-	-	**-.18**[b]	-	-.04	-
Overestimation	-.05[b]	-	**-.15**[b]	-	-	-	-	-	-	-	**-.09**[b]	-	**-.28**	-
Overplacement	.05	-	-.07[b]	-	-	-	-	-	-	-	**-.14**[b]	-	.00	-
Trivia	**.13**	.07	.10	.09	**.13**[a]	.01[b]	.05	.08	**.12**	**.12**	-	-	.07	**.11**
Confidence	-.02	.04	-.11[b]	-.01	.02	-.04	-.05	-.00	-.05	.02	-	-	-.03	.01
Overestimation	-.07	-.00	**-.18**[a]	-.03	-.07	-.05	-.06	-.06	**-.13**	-.07	-	-	**-.13**	-.09
Overplacement	.00[b]	.04[b]	-.07	.02	-.02	-.05	-.07	-.01	-.08[a]	.03[b]	-	-	-.07	-.08
CRT	**.10**[b]	.04[b]	**.17**[b]	.08[b]	.07[a,b]	-.07	**.15**[a,b]	.01[b]	**.14**[b]	.06[b]	.07[b]	-	**-.12**	-.09
Overclaiming												-		
Accuracy (*d'*)	**.20**	**.19**	**.30**[a]	**.20**[b]	**.26**[a]	-.01[b]	**.41**[a,b]	.07	**.40**[a,b]	**.13**	.02[b]	-	**.25**	.09
Bias (*c*)	.01[b]	.06	**.19**[a]	.06	.09[a,b]	-.08[b]	**.24**[a]	-.05[b]	**.29**[a]	.01	**.14**[b]	-	**.24**	.08
BSR												-		
Accuracy (*d'*)	**.18**	**.15**	**.24**	**.15**	**.12**[a]	-.03	**.17**	.12	**.20**[a]	**.17**[b]	-	-	**.18**	.03
Bias (*c*)	-.03[b]	.01	.09[b]	.06	-.04[b]	-.12[b]	.04[b]	-.03	**.16**[a]	.04	-	-	**.23**	-.01

Bolded correlations are correlations that were significant at the $p < .05$ threshold following Hommel-Bonferroni correction. The full results (pre-correction) are in Supplemental Table 10. *Ns* ranged from 498 to 1,502. Com. = Online community sample; GIHS = General Intellectual Humility Scale; CIHS = Comprehensive Intellectual Humility Scale; PIHS = Porter Intellectual Humility Scale; MIHS = McElroy Intellectual Humility Scale; AIHS = Alfano Intellectual Humility Scale; SIHS = Politics-Specific Intellectual Humility.

[a]Denotes correlations that were significantly different between samples according to tests of the difference between independent correlations.

[b]Denotes correlations that were significantly different from Honesty-Humility according to tests of the difference between dependent correlations.

that the relations between IH and critical thinking were larger in the online community sample than in the college sample.

Correlations between IH and (a) overconfidence and (b) overclaiming

Online community sample

By and large, IH was not related to indices of overconfidence (Table 3). Of 42 correlations, only 11 (26%) were significant, and most of these results were small. Although these relations often were small and/or not significant, they tended to be in a consistent direction, with IH being correlated with less overconfidence. As with critical thinking measures, IH measures that assessed primarily metacognitive features were largely similar in their relations with overconfidence and overclaiming to IH measures that assessed a blend of metacognitive, interpersonal, and emotional features. That said, the CIHS appeared to be a particularly consistent correlate of less overestimation compared with other IH measures.

Generally, IH measures were most consistently related to less overestimation across critical thinking tasks (7 of 13 correlations were significant, 54%), with the significant relations ranging from small to medium (significant *rs* ranged from -.09 [SIHS & political knowledge] to -.24 [CIHS & intelligence]). Regarding confidence and overplacement ratings, IH tended to not be significantly related to these ratings. The exceptions were the relations between the (a) CIHS and the confidence and overplacement ratings on the intelligence measure (*rs* were both -.09) and (b) the SIHS and the confidence and overplacement ratings on the political knowledge measure (*rs* were -.18 and -.14).

Whereas there was limited evidence that IH was related to less overconfidence, IH was consistently related to more accuracy (*d'*) on the Overclaiming Questionnaire (i.e. enhanced ability to discriminate between foil and real items), and the correlations were small to large (*rs* ranged from .02 [SIHS] to .41 [MIHS]). Moreover, IH tended to be weakly to moderately and positively related to response bias on the Overclaiming Questionnaire (i.e. tendency to endorse less familiarity as opposed to more familiarity across items; *rs* ranged from .01 [GIHS] to .29 [AIHS]).

College sample

In contrast with the online community sample, IH was not significantly related to any index of overconfidence (Table 3). Effect sizes were all small, and they were not in a consistent direction (e.g. some correlations were positive – such as the CIHS and overplacement on the ICAR – and some were negative – such as the PIHS and overplacement on the ICAR). IH measures that assessed primarily metacognitive features were largely similar in their relations with overconfidence and overclaiming to IH measures that assessed a blend of metacognitive, interpersonal, and emotional features. That said, the PIHS was, again, the only domain-general measure of IH to be negatively correlated with an indicator of accuracy, namely on the Overclaiming Questionnaire. Whereas IH was not related to measures of overconfidence in the college sample, IH tended to be related to more accuracy (*d'*) on the Overclaiming Questionnaire; these correlations were small to moderate (*rs* ranged from -.01 [PIHS] to .20 [CIHS]). IH was not significantly related to response bias on the Overclaiming Questionnaire.

Sample differences

Regarding sample differences, 20 of 40 correlations were significantly different (50% of the results, $ps < .05$) between the online community and college samples. Of these 20 significant differences, all results indicated that the relations between IH and overconfidence and overclaiming were larger in the online community sample than in the college sample.

Specificity to IH

The correlations between IH measures and Honesty-Humility ranged from .04 (SIHS) to .49 (AIHS; Supplemental Table S1). To examine whether the relations were specific to IH, we conducted tests of the difference between dependent correlations using Fisher's *r*-to-*z* transformation and interpreted results based on two-tailed *p*-values (Lee & Preacher, 2013). Of 144 analyses across samples, only 49 (34%) indicated that the differences in the relations between IH and critical thinking and overconfidence and Honesty-Humility and critical thinking and overconfidence were significant (see Table 3). Of these 49 significant results, 25 (51%) revealed that the relations between Honesty-Humility and critical thinking and overconfidence were significantly larger than the correlations between IH and these measures. The remaining 24 (49%) revealed that the relations between IH and critical thinking and overconfidence were significantly larger than the correlations between Honesty-Humility and these measures.

There were also differences across measures. Regarding outcome measures, IH tended to be

a stronger correlate of more critical thinking than Honesty-Humility whereas Honesty-Humility tended to be a stronger correlate of (a) less overconfidence and (b) a greater tendency to view items as not profound on the bullshit receptivity measure than IH. Regarding IH measures, for the PIHS, GIHS, and MIHS the majority of the significant differences suggested that the relations were stronger for Honesty-Humility than IH in both samples (15 of 19 correlations across measures). Similarly, most significant differences indicated that correlations were larger for Honesty-Humility than the SIHS (4 of 7 correlations). In contrast, for the CIHS and AIHS, most of the significant differences suggested that the correlations were larger for the CIHS (4 of 5 correlations) and the AIHS (5 of 6 correlations) than for Honesty-Humility.

Is there evidence for a Dunning-Kruger effect across IH quartiles?

The distribution of perceived vs. actual scores for the intelligence, political knowledge, and trivia knowledge measures are reported in Figures 1–3. The *t*-test statistics are reported in online supplemental materials.

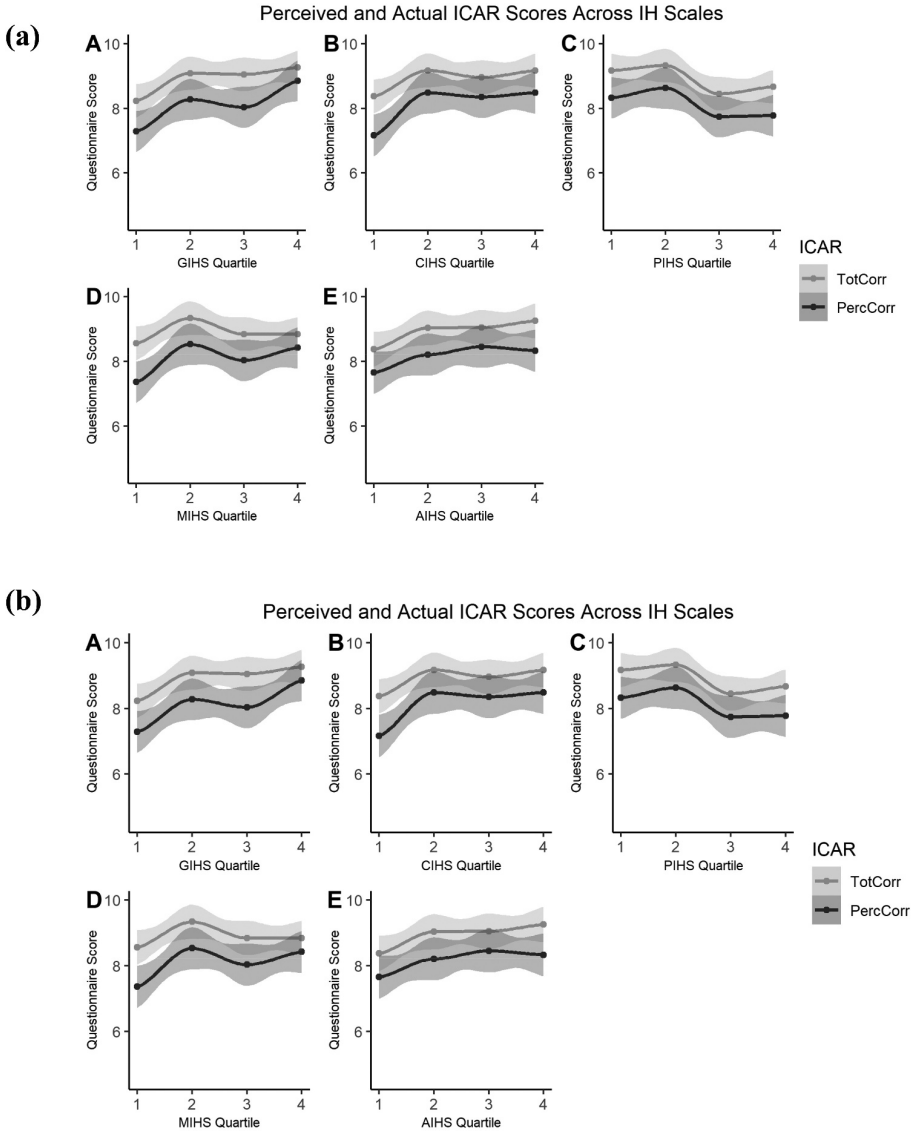

Figure 1. Objective accuracy scores vs. perceived accuracy scores on the ICAR in (a) online community samples and (b) the college sample across IH measures. GIHS = General Intellectual Humility Scale; CIHS = Comprehensive Intellectual Humility Scale; PIHS = Porter Intellectual Humility Scale; MIHS = McElroy Intellectual Humility Scale; AIHS = Alfano Intellectual Humility Scale; SIHS = Politics-Specific Intellectual Humility; ICAR = Intelligence; TotCorr = Total Correct, PercCorr = Perceived Correct

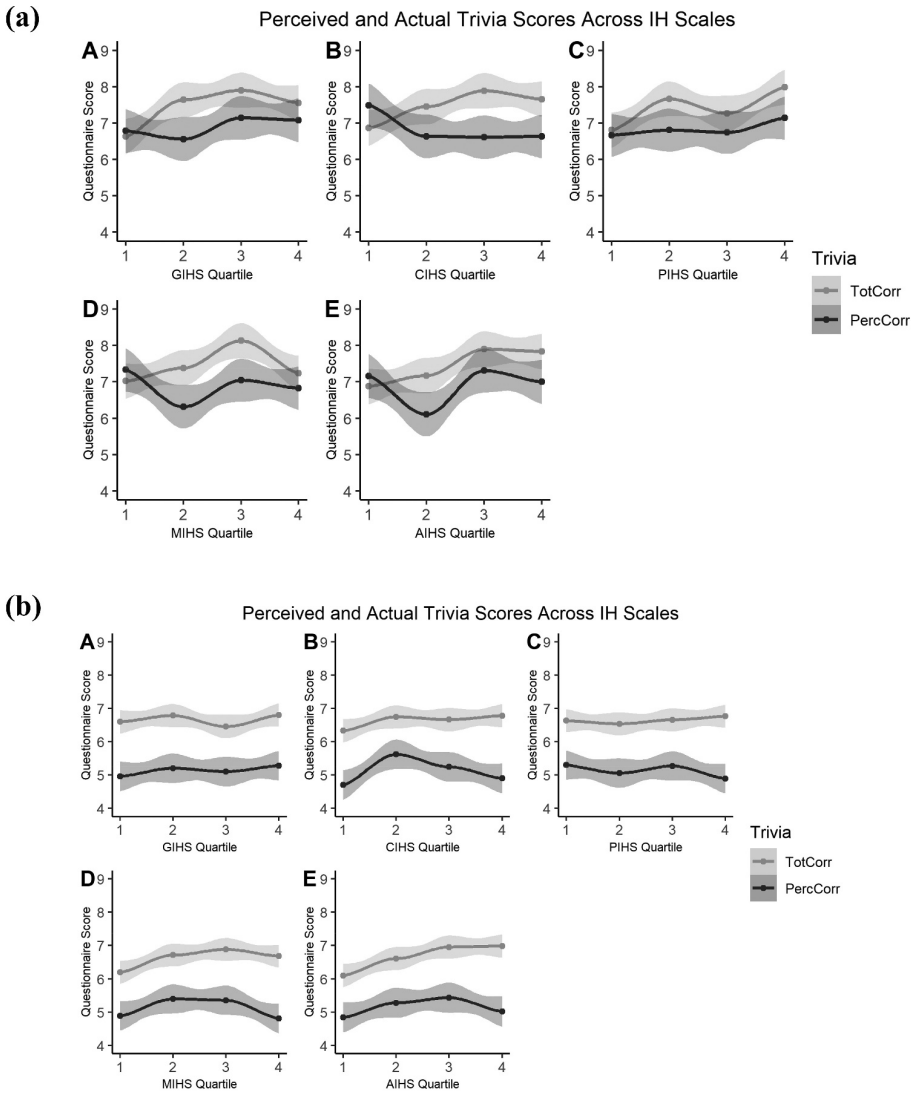

Figure 2. Objective accuracy scores vs. perceived accuracy scores on the trivia knowledge scale in (a) online community samples and (b) the college sample across IH measures. GIHS = General Intellectual Humility Scale; CIHS = Comprehensive Intellectual Humility Scale; PIHS = Porter Intellectual Humility Scale; MIHS = McElroy Intellectual Humility Scale; AIHS = Alfano Intellectual Humility Scale; SIHS = Politics-Specific Intellectual Humility; TotCorr = Total Correct, PercCorr = Perceived Correct

Online community sample

Overall, there was some evidence for a DK effect in the online community sample. There was consistent evidence for a DK effect on the intelligence measure and political knowledge measure. That is, participants scoring at the 1st quartile on all IH measures significantly overestimated their scores on the intelligence and political knowledge measures compared with their actual performance on these tasks (*ps* < .05, *ds* ranged from .12 to .59).[4] On the trivia measure, however, there was scant evidence for a DK effect. Indeed, there was only evidence for a DK effect at the 1st quartile of the CIHS ($p < .05$, $d = .18$). What is more, there was evidence for an underestimation effect on the trivia knowledge measure, such that at moderate to high levels of IH across all IH measures (2nd to 4th quartiles), participants significantly underestimated their performance relative to their actual performance (*ps* < .05, *ds* ranged from .20 to .42).

College sample

In contrast to the online community sample, there was no evidence for a DK effect in the college sample. Instead, there was consistent evidence that participants underestimated their performance irrespective of their level of IH. In other words, whether scoring at the 1st or 4th quartile on IH measures, participants underestimated their performance on all critical thinking tasks relative to

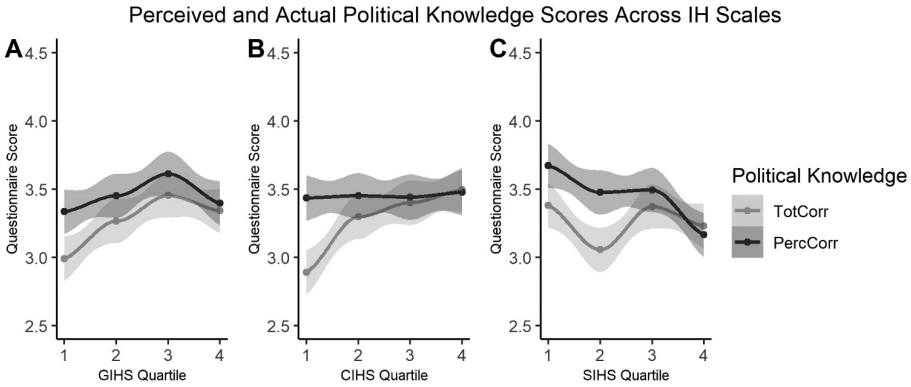

Figure 3. Objective accuracy scores vs. perceived accuracy scores on the political knowledge scale in online community samples across IH measures. GIHS = General Intellectual Humility Scale; CIHS = Comprehensive Intellectual Humility Scale; PIHS = Porter Intellectual Humility Scale; MIHS = McElroy Intellectual Humility Scale; AIHS = Alfano Intellectual Humility Scale; SIHS = Politics-Specific Intellectual Humility; TotCorr = Total Correct, PercCorr = Perceived Correct

their actual performance (*ps* < .05, *ds* ranged from .13 to .82). This was particularly the case on the trivia knowledge measure, as participants significantly underestimated their performance on the trivia knowledge measure across all quartiles of IH across all measures of IH.

Discussion

The current work, which involved online community and college participants and included multiple measures of IH, critical thinking, and accuracy, offers new insights onto the relations between IH and accuracy and overconfidence. We found that IH was generally related to more critical thinking, less overestimation, and less overclaiming. Moreover, there was evidence that those scoring the lowest on IH significantly overestimated their performance on intelligence and political knowledge measures, thus providing some evidence of a Dunning-Kruger effect. Importantly, the Dunning-Kruger effect was only present in the community sample; the DK effect did not appear to generalize to the college sample. Below we describe our findings in greater detail as we look forward to additional research on this topic.

Are intellectually humble individuals more accurate?

IH tended to be a small-to-medium correlate of more critical thinking, including more intelligence, trivia knowledge, cognitive reflection, political knowledge, and abilities to distinguish between semantically meaningful and meaningless statements. These results are in line with previous research indicating that IH is related to more critical thinking (e.g. Zmigrod et al., 2019), less biased decision-making (e.g. Bowes et al., 2022), less misinformation susceptibility (Bowes & Tasimi, 2022), and more interest in learning from one's mistakes (Porter et al., 2020).

In both samples, correlations with accuracy tended to be consistent across IH conceptualizations. Given that the majority of critical thinking measures we used were strictly knowledge-based in the sense that they did not ask about opinions, beliefs, and attitudes, it stands to reason that the metacognitive aspects of IH would be equally strong contributors to more accuracy as the interpersonal and emotional aspects of IH. Note that these results contrast with previous research indicating that IH measures differ in their relations with misinformation susceptibility and conspiratorial ideation (e.g. Bowes et al., 2020; Bowes & Tasimi, 2022); yet, misinformation susceptibility and conspiratorial ideation reflect opinions, beliefs, and attitudes, all of which are heavily imbued with interpersonal (e.g. ingroup vs. outgroup) and emotional features (e.g. cognitive dissonance). Although our results do not address the question of whether different IH features are relevant for different forms of knowledge, reasoning, and beliefs, future research should continue to examine these potential differences. For instance, IH might be related to 'cold' reasoning (i.e. knowledge-based, logical reasoning) across IH features whereas the interpersonal and emotional features of IH might be especially important for 'hot' reasoning (i.e. emotional, identity-relevant reasoning).

Are intellectually humble individuals less overconfident?

From our point of view, the answer to this question is: it depends. The relations between IH and overconfidence were variable across critical thinking measures. In general, there was little evidence that IH was related to confidence and overplacement ratings on the critical thinking measures. There was, however, evidence that IH was weakly related to less overestimation on the critical thinking measures, meaning intellectually humble individuals were slightly more calibrated in their assessment of their performance relative to their actual performance. IH was also consistently related to more accuracy on the overclaiming measure, and relations tended to be medium to large. The overconfidence measures pertained to reflecting on one's level of problem-solving accuracy whereas the overclaiming measure pertained to reflecting on whether one had heard of a particular topic or not. Thus, IH may be a stronger and more consistent correlate of the ability to self-reflect regarding one's *familiarity* with topics as opposed to one's *performance* on critical thinking tasks.

Relations between IH and overconfidence and overclaiming were largely similar across IH conceptualizations. These results were consistent with the relations between IH and critical thinking and demonstrate that there are few differences across narrow and broad measures of domain-general IH. Although there were few differences across IH measures, the CIHS emerged as the most consistent predictor of less overconfidence. Recall that the CIHS includes metacognitive, interpersonal, and emotional features, all of which are also involved in judgments of accuracy. Perhaps this result may not be so surprising given that one of the subdimensions of the CIHS is Lack of Intellectual Overconfidence (see Supplemental Materials 4). In fact, in exploratory and subsidiary analyses, Lack of Intellectual Overconfidence was a medium-to-large negative correlate of all measures of overconfidence across critical thinking tasks (confidence, overestimation, overplacement; see Supplemental Tables S6-S7). Future research should probe into these features of IH, specifically those that pertain to recognizing that one can change their beliefs even when they are confident, turn to others for their expertise, and listen to others' perspectives, in the context of knowledge self-assessment.

Are relations specific to IH?

Of 144 analyses across samples, only one-third of the results indicated a significant difference between Honesty-Humility and IH in their relations with critical thinking and overconfidence. Of these significant correlations, a majority of the results indicated that the relations were larger for Honesty-Humility than for IH. Yet, there were important nuances in these results. First, Honesty-Humility was a stronger correlate of less overconfidence and a greater tendency to view items as not profound on the bullshit receptivity measure than IH. In contrast, IH was a significantly stronger correlate of more critical thinking than Honesty-Humility. Hence, relations with critical thinking appear to be specific to IH relative to Honesty-Humility whereas that does not seem to be the case for overconfidence.

Nevertheless, for the two multidimensional IH measures that assess intrapersonal, interpersonal, and emotional IH features (the CIHS and AIHS), most of the correlations were stronger for IH than for Honesty-Humility. These results raise the possibility that a combination of IH features is a unique or specific predictor of critical thinking and overconfidence relative to general humility features. Thus, whereas there were few differences between IH conceptualizations relative to *each other* in their relations with critical thinking and overconfidence, there are differences across IH conceptualizations relative to *general humility features* in their relations with critical thinking and overconfidence. These results in aggregate provide additional evidence for the construct validity of multidimensional IH measures.

Is there a Dunning-Kruger effect for intellectual humility?

There was consistent evidence for a Dunning-Kruger (DK) effect on the intelligence and political knowledge measures in the online community sample. Those scoring at the 1st quartile of IH significantly overestimated their performance relative to their actual performance, meaning participants who were the least intellectually humble overestimated their performance on the intelligence and political knowledge measures the most.

Our results revealed that the trivia knowledge measure functioned differently from the intelligence and political knowledge measures in relation to IH. Across both samples, participants tended to underestimate their performance on the trivia knowledge measure relative to their actual performance. This underestimation manifested across IH quartiles, meaning that it was largely irrespective of IH levels. These results suggest that IH does not necessarily influence one's ability to evaluate their performance on the trivia knowledge task whereas IH does seem to influence this ability on the intelligence and political knowledge tasks (at least in the online community sample).

Previous research has found that if a task is perceived as 'too difficult' or is objectively difficult, including a trivia knowledge task, participants cannot accurately assess their performance on the task. This difficulty persists across levels of performance such that those who are both 'skilled' and 'unskilled' become inaccurate at assessing their performances on hard tasks (Burson et al., 2006; see also Dunning, 2011, for a review). In the present investigation, the average score on the trivia knowledge measure indicated that participants gave the correct response on less than half the items in both samples, raising the possibility that the task was too difficult. This difficulty may have washed out not only the possibility for a DK effect but also room for IH to make a difference in self-evaluation.

Sample differences in the relations between IH and accuracy and overconfidence

There were several differences between the online community and college samples. First, approximately one-third of the correlations between IH and critical thinking were significantly larger in the online community sample than in the college sample. Similarly, half of the correlations between IH and overconfidence and overclaiming were significantly larger in the online community sample than in the college sample. What is more, there was no evidence for a DK effect in the college sample. Across IH quartiles, undergraduate participants underestimated their performance on all critical-thinking tasks relative to their actual performance.

What might explain these differences between the two samples? For the critical thinking and overconfidence results, secondary analyses revealed that neither age nor education were sufficient in explaining these differences. In the online community sample, age tended to not significantly moderate the relations between IH and critical thinking. Education was also not a sufficient explanation for these differences across samples, as IH tended to remain a significant predictor of critical thinking and overconfidence even when controlling for education. These findings are consistent with other research that has found that demographic characteristics, such as age and education, do not explain differences in personality traits and values between online community and college samples (Weigold & Weigold, 2021).

Identifying the factors that may explain differences between community and college samples has been a challenging endeavor for decades (see Peterson, 2001); for instance, even cultural worldviews have been found not to explain differences between community and college samples in studies from dozens of countries (e.g. Hanel & Vione, 2016). Adding to the murkiness of classifying sample differences, some research shows that relations among psychological variables tend to be stronger in college samples (e.g. Peterson, 2001) whereas other research shows that relations among psychological variables tend to be stronger in online community samples (e.g. Weigold & Weigold, 2021). In many ways, this variability makes sense, as there is as much heterogeneity within college samples as there is within online community samples (Hanel & Vione, 2016). College and online community samples also differ on a multitude of demographic variables, from socioeconomic status to relationship status to sexual identity to racial identity (e.g. Weigold & Weigold, 2021). Thus, as has been the case for prior work, why sample differences emerge is a question deserving of additional research.

That said, we think it is worth considering two possibilities that might shed light on the sample differences reported here. The first is developmental in nature – that is, adolescents and young adults have a less stable sense-of-self than adults and are going through a number of crucial transitions pertaining to identity development (e.g. moving out, forming new interpersonal relationships; see Gallander Wintre et al., 2001; Sears, 1986). The overconfidence and overclaiming results directly stem from self-reflection processes that involve thinking about one's own performance and the performance of others. Because young adults have a more 'in flux' sense-of-self than older adults and still developing self-reflection abilities, it is possible that these developmental differences at least partially explain the observed sample differences.

Consistent with the idea of shifting identities in young adulthood is the idea of shifting peer groups in young adulthood. Another potential interpretation of these results is from the perspective of social comparison – undergraduate students likely compare their performance to other undergraduate students. They may perceive that others are equally or more capable on a given task, which contributes to inaccurate underestimation at the higher end of performance (see Dunning, 2011). Similarly, it is possible this null effect arises from the effect of a classroom environment rather than level of education or age cohort. The undergraduate environment may constrain self-evaluation, particularly in the absence of feedback. Thus, there may not be room for IH to influence self-evaluations on specific critical-thinking tasks in undergraduate students. Of course, research is needed to test these possibilities.

Limitations and future directions

The present investigation is characterized by a number of strengths that distinguish it from other studies in this domain, such as our inclusion of (a) multiple measures of IH, critical thinking, and overconfidence, and (b) both online community and college participants. Our study was arguably the most comprehensive investigation of the relations between IH and critical thinking and overconfidence to date. Nevertheless, there are limitations that warrant consideration.

First, the relations in the present study were generally small. Given that definitions of IH are largely centered on a greater ability to self-reflect, the small effects found in our study may raise questions surrounding conceptualizations of IH. That said, correlations between self-report measures and behavioral measures tend to be small-to-medium and are rarely large, due at least in part to the noisiness and low reliabilities often inherent in behavioral tasks (e.g. Dang et al., 2020). Thus, it is not altogether surprising that the relations between IH and critical thinking are small purely for largely unavoidable methodological reasons. Moreover, there was evidence for floor effects on the overconfidence measures. Participants, in general, *underestimated* rather than *overestimated* their performance on critical thinking tasks. Thus, there was restriction of range that could have attenuated the relations between IH and overconfidence.

In addition, even though the cross-sectional relations are small, it is possible that, when measured over time, the relations are stronger. To elaborate, IH may confer more accuracy; more accuracy may, in turn, reinforce IH and seeking out new information, thus compounding the strength of the relations over time. In other words, the temporal and causal relations between IH and critical thinking and overconfidence may be stronger than the cross-sectional relations, a possibility that has not yet been examined.

Just as it will be important to examine how these relations unfold over time, it will also be important to examine these relations at different ages. If IH is related to more critical thinking and accurate self-assessment in children and adolescents, then IH is an especially promising target for applied research aiming to increase IH. Some research indicates that IH in children (6–8 years old) is related to more intelligence and event-related potential markers of paying attention to one's mistakes (Danovitch et al., 2019). By increasing IH earlier in life, it may be possible to increase its benefits, chief among them being able to acquire information more effectively and understand the limits of one's abilities.

Moreover, future research is needed to continue disentangling the critical thinking results from the overconfidence results. It is clear from the results of this study, in addition to the broader literature on accuracy and overconfidence, that the processes contributing to *accuracy* are not isomorphic to the processes contributing to *judgments of accuracy* (see Dunning, 2011). Whereas accuracy pertains to primarily metacognitive processes (e.g. memory, knowledge), accuracy judgments pertain to metacognitive, interpersonal (e.g. 'how did the average person perform relative to me?'), and emotional (e.g. 'will I feel bad about myself if I did poorly on this task?') processes (see Moore & Healy, 2008). There are motivational pulls to uphold one's sense-of-self and mitigate negative emotions that play into making judgments of one's performance (see Dunning, 2011). Considering these differences, the mechanisms by which IH contributes to more accuracy are likely separable from the mechanisms by which IH contributes to less overconfidence. Research is needed to clarify why and how accuracy and overconfidence overlap and diverge.

Conclusion

Our findings demonstrate that IH is related to more accuracy and less overestimation and overclaiming. What this means is that IH is not only relevant to how people solve problems but also to their perceptions of how they solve such problems. Results indicate that there were differences between online community samples and college samples in the context of IH, critical thinking, and accuracy. These differences invite important questions about the roles of social context in the ability to self-reflect on one's abilities. Our results also suggest that certain measures of IH are stronger correlates of critical thinking and overconfidence than general humility, specifically those that capture multiple dimensions of IH features. Along these lines, certain IH characteristics, such as lacking intellectual overconfidence, may be more powerful than others in this context. Thus, IH may contribute to more accurate and balanced decision-making, and future research can clarify whether and how our findings generalize to real-world contexts where the stakes might be high.

Notes

1. Due to a computer error, one item was left out of the AIHS and the subdimension of Modesty: 'I wouldn't want people to treat me as though I were intellectually superior to them'.
2. Due to an error, one of the 'real' items was not included in the survey battery in Sample 3 (31 'real' and 13 'foil' items).
3. To further probe potential differences across samples, we conducted three subsidiary analyses. First, we examined age as a moderator in the relations between IH and critical thinking and overconfidence in the online community sample. Of 79 analyses, 18 (23%) were significant. Of these significant results, all indicated that the relations between IH and critical thinking and overconfidence were stronger in younger than in older participants. We also covaried for education in the relations between IH and critical thinking and overconfidence in the community sample. IH accounted for an average 1–2% of the variance in critical thinking and overconfidence over-and-above education. Finally, we examined the correlations between IH and critical thinking and overconfidence in college-aged participants in the online community sample (age of 26 or younger). The correlations in this subsample were larger than the correlations in the student sample. Put differently, the results broadly mirrored the correlations in the larger online community sample rather than the college sample (Supplemental Table 9). All results are available in online supplemental materials.
4. Consistent with Leman et al. (2021), we also examined overestimation using a one-way ANOVA and follow-up Bonferroni-adjusted contrasts. These results are available in Supplemental Materials 5.

Acknowledgments

We would like to thank the late Dr. Scott O. Lilienfeld for his contributions to this project.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability

The first-author was responsible for study conceptualization, data collection, data preparation, data analyses, and report writing. The second-author was responsible for data preparation and data analyses. The third-author was responsible for study conceptualization and report writing.

Open scholarship

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The data and materials are openly accessible at https://doi.org/10.1080/17439760.2023.2208100

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., Lapsley, D., & Tractenberg, R. E. (2017). Development and validation of a multi-dimensional measure of intellectual humility. *PloS One*, *12*(8), e0182950. https://doi.org/10.1371/journal.pone.0182950

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, *103*(4), 718–735. https://doi.org/10.1037/a0029395

Anson, I. G. (2018). Partisanship, political knowledge, and the Dunning-Kruger effect. *Political Psychology*, *39*(5), 1173–1192. https://doi.org/10.1111/pops.12490

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498. https://doi.org/10.1037/0033-2909.110.3.486

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*(5), 2–23. https://doi.org/10.1016/j.amjmed.2008.01.001

Bowes, S. M., Blanchard, M. C., Costello, T. H., Abramowitz, A. I., & Lilienfeld, S. O. (2020). Intellectual humility and between-party animus: Implications for affective polarization in two community samples. *Journal of Research in Personality*, *88*, 103992.

Bowes, S. M., Costello, T. H., Lee, C., McElroy-Heltzel, S., Davis, D. E., & Lilienfeld, S. O. (2022). Stepping outside the echo chamber: Is intellectual humility associated with less political myside bias? *Personality & Social Psychology Bulletin*, *48*(1), 150–164. https://doi.org/10.1177/0146167221997619

Bowes, S. M., & Tasimi, A. (2022). Clarifying the relations between intellectual humility and pseudoscience beliefs, conspiratorial ideation, and susceptibility to fake news. *Journal of Research in Personality*, *98*, 104220. Advance online publication. https://doi.org/10.1016/j.jrp.2022.104220

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*(1), 60–77. https://doi.org/10.1037/0022-3514.90.1.60

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, *64*(3), 391–418.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

Danovitch, J. H., Fisher, M., Schroder, H., Hambrick, D. Z., & Moser, J. (2019). Intelligence and neurophysiological markers of error monitoring relate to children's intellectual humility. *Child Development*, *90*(3), 924–939. https://doi.org/10.1111/cdev.12960

Deffler, S. A., Leary, M. R., & Hoyle, R. H. (2016). Knowing what you know: Intellectual humility and judgments of recognition memory. *Personality and Individual Differences*, *96*, 255–259. https://doi.org/10.1016/j.paid.2016.03.016

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In James M. Olson, & Mark P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). Academic Press.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gallander Wintre, M., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology/Psychologie Canadienne*, *42*(3), 216–225. https://doi.org/10.1037/h0086893

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

Hanel, P. H., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PloS One*, *11*(12), e0168354.

Hoyle, R. H., Davisson, E. K., Diebels, K. J., & Leary, M. R. (2016). Holding specific views with humility: Conceptualization and measurement of specific intellectual humility. *Personality and Individual Differences*, *97*, 165–172. https://doi.org/10.1016/j.paid.2016.03.043

Jenkins, M. M., & Youngstrom, E. A. (2016). A randomized controlled trial of cognitive debiasing improves assessment and treatment selection for pediatric bipolar disorder. *Journal of Consulting and Clinical Psychology*, *84*(4), 323–333. https://doi.org/10.1037/ccp0000070

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121. https://doi.org/10.1037/0022-3514.77.6.1121

Krumrei-Mancuso, E. J., Haggard, M. C., LaBouff, J. P., & Rowatt, W. C. (2020). Links between intellectual humility and acquiring knowledge. *The Journal of Positive Psychology*, *15*(2), 155–170. https://doi.org/10.1080/17439760.2019.1579359

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. https://doi.org/10.1080/00223891.2015.1068174

Larrick, R. P. (2004). Debiasing. In D. J. Koehler, N. Harvey (Eds). *Blackwell Handbook of Judgment and Decision Making* (pp. 316–338). Blackwell Publishing.

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality & Social Psychology Bulletin*, *43*(6), 793–813. https://doi.org/10.1177/0146167217697695

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Lee, I. A., & Preacher, K. J. (2013). *Calculation for the test of the difference between two dependent correlations with one variable in common* [computer software]. http://quantpsy.org/corrtest/corrtest2.htm

Leman, J., Kurinec, C., & Rowatt, W. (2021). Overconfident and unaware: Intellectual humility and the calibration of metacognition. *The Journal of Positive Psychology*, *18*(1), 1–19. https://doi.org/10.1080/17439760.2021.1975155

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington, E. L., Jr., & Van Tongeren, D. R. (2014). Intellectual humility: Scale development and theoretical elaborations in the context of religious leadership. *Journal of Psychology and Theology*, *42*(1), 19–30. https://doi.org/10.1177/009164711404200103

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, *11*(8), 1–12. https://doi.org/10.1111/spc3.12331

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*(4), 890. https://doi.org/10.1037/0022-3514.84.4.890

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563. https://doi.org/10.1017/S1930297500006999

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *The Journal of Consumer Research*, *28*(3), 250–261. https://doi.org/10.1086/323732

Porter, T., & Schumann, K. (2018). Intellectual humility and openness to the opposing view. *Self and Identity*, *17*(2), 139–162. https://doi.org/10.1080/15298868.2017.1361861

Porter, T., Schumann, K., Selmeczy, D., & Trzesniewski, K. (2020). Intellectual humility predicts mastery behaviors when learning. *Learning and Individual Differences*, *80*, 101888. https://doi.org/10.1016/j.lindif.2020.101888

Preacher, K. J. (2002, May). Calculation for the test of the difference between two independent correlation coefficients [Computer software]. Available from http://quantpsy.org.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273–1296.

Teo, N. Q., Lim, T., & Tong, E. M. (2022). The humble estimate: Humility predicts higher self-assessment accuracy. *British Journal of Social Psychology*, *62*(1), 561–582. https://doi.org/10.1111/bjso.12545

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Weigold, A., & Weigold, I. K. (2021). Traditional and modern convenience samples: An investigation of college student, mechanical turk, and mechanical turk college student samples. *Social Science Computer Review*, *40*(5), 1302–1322. Advanced online publication. https://doi.org/10.1177/08944393211006847

Zmigrod, L., Zmigrod, S., Rentfrow, P. J., & Robbins, T. W. (2019). The psychological roots of intellectual humility: The role of intelligence and cognitive flexibility. *Personality and Individual Differences*, *141*, 200–208. https://doi.org/10.1016/j.paid.2019.01.016