1-7-2022

The Validity of the General Intellectual Humility Scale as a Measure of Intellectual Humility

Charles Westbrook

ACCEPTANCE

This dissertation, THE VALIDITY OF THE GENERAL INTELLECTUAL HUMILITY SCALE AS A MEASURE OF INTELLECTUAL HUMILITY, by CHARLES WESTBROOK, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education & Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

_____
Don Davis, Ph.D.
Committee Chair

_____
Jeff Ashby, Ph.D.
Committee Member

_____
Maggie Renken, Ph.D.
Committee Member

_____
Ken Rice, Ph.D.
Committee Member

_____
Date

_____
Brian Dew, Ph.D.
Chairperson, Department of Counseling and
Psychological Services

_____
Paul A. Alberto, Ph.D.
Dean, College of Education &
Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education & Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

_____

Charles Westbrook

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Charles Joseph Westbrook
Counseling and Psychological Services
College of Education & Human Development
Georgia State University

The director of this dissertation is:

Don Davis
Department of Counseling and Psychological Services
College of Education & Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Charles Westbrook

EDUCATION:

Ph.D. in Counseling Psychology	2014-Present	Georgia State University Department of Counseling and Psychological Services
M.S. in Clinical Mental Health Counseling	2014	Georgia State University Department of Counseling and Psychological Services
B.A.	2005	University of Georgia Philosophy

PROFESSIONAL EXPERIENCE:

2019-present	Psychotherapist David E. Loftis, PhD
2018-2019	Doctoral Intern Georgia State University Counseling & Testing Center
2012-2018	Graduate Assistant Georgia State University
2016-2017	Intern; Psychological Assistant Georgia Pediatric Psychology
2016-2017	Practicum Student Oglethorpe University Counseling Center
2015-2016	Practicum Student Georgia Institute of Technology Counseling Center

2014-2015	Practicum Student Georgia State University Career Services
2013-2014	Intern Skyland Trail

PRESENTATIONS AND PUBLICATIONS:

Westbrook, C. J., Davis, D. E., McElroy, S.E., Brubaker, K, Choe, E, Karaga, S., Dooley, M., O'Bryant, B. L., Van Tongeren, D., Hook, J. (in press). Trait Sources of Spirituality Scale: Assessing Trait Spirituality More Inclusively. *Measurement and Evaluation in Counseling and Development*.

McElroy, S. E., Choe, E., **Westbrook, C. J**., Davis, D. E., Hook, J. N., Van Tongeren, D. (in press). Relational Spirituality and Forgiveness of Intergroup Offenses. *Journal of Theology and Psychology*.

Davis, D. E., McElroy, S. E., Rice, K. G., Choe, E., **Westbrook, C. J.**, Hook, J. N., Van Tongeren, D., DeBlaere, C., Hill, P., Placeres, V., Worthington Jr, E. L. (2016). Is Modesty a Subdomain of Humility. *The Journal of Positive Psychology*, *11*(4), 439-446.

Choe, E., Davis, D.E., McElroy, S., **Westbrook, C.J.**, van Nuenen, M., Van Tongeren, D.R., & Hook, J.N. (2016). Relational Spirituality and Forgiveness of Offenses Committed by Religious Leaders. *The International Journal of the Psychology of Religion*, *26*(1), 46-60.

Davis, D. E., Ho, M., Griffen, B., Bell, C., Hook, J., Van Tongeren, D., DeBlaere, C., Worthington, E., Jr., **Westbrook, C. J.** (2015). Forgiving the Self and Physical and Mental Health Correlates: A Meta-analytic Review. *Journal of Counseling Psychology*, *62*(2), 329.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2014-2018	Counseling Psychology Student Organization | Member
2015-2016	American Psychological Association (Div. 5, 13, 17, 24) | Student Affiliate
2015-2016	Association for Psychological Science | Student Affiliate
2015-2016	Georgia Psychological Association | Student Affiliate
2013-2014	Licensed Professional Counselors Association of Georgia | Student Affiliate
2013-2014	Chi Sigma Iota
2012-2014	American Counseling Association

The Validity of the General Intellectual Humility Scale as a Measure of Intellectual Humility

by

Charles Westbrook

Under the Direction of Don Davis, Ph.D.

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Education and Human Development

Georgia State University

2021

ABSTRACT

Early intellectual humility research has largely relied on questionnaires that require individuals to self-evaluate their own intellectual humility, despite concerns that people low in intellectual humility may lack awareness of their degree of intellectual humility. Because of this potential source of error, it is important that self-report measures of intellectual humility are thoroughly tested for validity. In Chapter 1, I conducted a systematic literature review of measures of intellectual humility. For each measure, validity evidence is summarized and critically evaluated. Validity evidence was found lacking with respect to addressing potentially serious problems with self-report. This finding points to a need for additional validity testing for self-report measures of intellectual humility. In Chapter 2, I conducted a set of pointed tests of validity for one such measure, the General Intellectual Humility Scale (GIHS). In a sample recruited from Prolific (N = 481), GIHS scores were weakly associated with or unassociated with endorsement of epistemically unwarranted beliefs, unassociated with endorsing such beliefs as certainly true, and unassociated with endorsing such beliefs despite claiming to have carefully researched the issue. Additionally, GIHS scores predicted greater bias blind spot, and this effect remained significant when controlling for science intelligence. Finally, GIHS scores predicted belief in anthropogenic global warming when controlling for political orientation but did not attenuate political polarization about global warming. I argue that these findings are clear departures from theory yet are consistent with suspected problems with direct self-report. I conclude by discussing limitations and implications for future research.

INDEX WORDS: General Intellectual Humility, Measurement, Validity

DEDICATION

This work is dedicated to my wife, Melissa. It has been a long and difficult journey for both of us, and I could not have made it through without your strength and your sacrifice. Especially over this final stretch during the pandemic, you have carried far more than your fair share so that we could reach the finish line. Thank you for believing in me even when I had trouble believing in myself. I am more impressed by you with each passing year. This work is also for my children, Simon, Ian, Arthur, and Heath. I know the hours I've had to put in for this project have been hard on you as well, and while I'm not going to be entirely done with work once I graduate (like you sometimes have suggested), I am excited to be spending more time with you soon. I am also deeply grateful for the support of my other family members, my friends, and the many others who helped me along the way. I could not have done it without you. Finally, this work is in memory of Dan Kemper, my father-in-law and one of the very best people I have ever known.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Michael Massengale for his invaluable feedback at every stage of this project and for his steady support. Additionally, I would like to think Dr. Kathleen Dyer and Dr. Ray Hall for their permission to use their measure and for their helpful guidance on its use in my project. I am grateful to Dr. Paul Barrett, Dr. Femke Truijens, Dr. Lee Branum-Martin, and Dr. Andrew Maul for their assistance and encouragement in my process of learning about measurement validity. I would also like to express my appreciation for the late Dr. Scott Lilienfeld for being an early inspiration of mine and for his support at a key juncture in my working on this project.

I would like to thank my committee members, Dr. Maggie Renken, Dr. Ken Rice, and Dr. Jeff Ashby. They have been formative in my thinking about psychology, consistently generous with their time, and have treated me with tremendous kindness over the years. I consider myself fortunate to have had the privilege of learning from each of them and have tremendous respect for all three of them as scholars and as people.

Finally, I am forever indebted to my advisor, Dr. Don Davis, for the opportunity to realize a decades-long dream in pursuing this degree and for his endless patience in helping me to this point. I look back now on our first years of working together with the benefit of hindsight and am struck by how was supportive he was of me pursuing my own interests even as he was still working to earn tenure. I will always be grateful for his deep commitment to helping me succeed and for his warmth and humanity throughout the process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 A CRITICAL REVIEW OF MEASURES OF TRAIT INTELLECTUAL HUMILITY

In the past decade, intellectual humility has emerged as a topic of interest in psychological research (VanTongeren et al., 2019). Whereas early work developing measures of general humility proceeded somewhat slowly (Davis et al., 2010), early intellectual humility research has resulted in the rapid development and publication of several measurement instruments in parallel (McElroy-Heltzel et al., 2019; Porter et al., 2021). With an abundance of simultaneously developed instruments now published, each featuring different definitions, conceptualizations, and decisions about validity testing, there is a need to evaluate how they are conceptually related to each other, the evidence supporting each instrument, and what additional validity evidence is needed.

Intellectual humility has its roots in an area of philosophy known as virtue epistemology (Baehr, 2011). Whereas virtue epistemology as a topic encompasses a diverse body of philosophical thought, generally virtue epistemology concerns the role of intellectual character in addressing foundational questions about knowledge and other forms of cognitive success (Alfano et al., 2019). Among the projects undertaken by virtue epistemologists are efforts to identify and understand specific epistemic virtues and vices, and intellectual humility has been noted as one of these virtues (Baehr, 2011; Roberts & Wood, 2003).

Several formal accounts of the nature of intellectual humility have been published in recent years, mostly in conjunction with the emergence of intellectual humility as a topic in psychology (Church, 2016; Hazlett, 2012; Roberts & Wood, 2003; Samuelson, et al., 2015; Tanesini, 2018; Whitcomb et al., 2017). Of these accounts, it is the limitations-owning account of Whitcomb et al. (2017) that has perhaps had the greatest influence on psychological measures

of intellectual humility. It has been acknowledged as influential or foundational in the conceptual work associated with several measures (Alfano et al., 2017; Haggard et al., 2018; Hoyle et al., 2016; Leary et al., 2017; Zachry et al., 2018), and awareness of limitations is the only content theme reflected in the items of every measure of intellectual humility (Porter et al., 2021, p. 5). The limitations-owning account is an attempt to provide a definition of intellectual humility that encompasses the strengths of prior conceptual work on intellectual humility while addressing problems and ambiguities of those accounts. On the limitations-owning account, intellectual humility is defined as the proper attentiveness to and ownership of one's intellectual limitations, which is generally described as a disposition to properly be aware of and take responsibility for intellectual limitations (Whitcomb et al., 2017). As an attempt at unifying prior work, it is argued that many qualities expected of intellectual humility follow from this either directly or as general psychological expectations, including the stronger elements of other accounts of intellectual humility – holding beliefs with appropriate firmness (Samuelson et al., 2015), having proper beliefs about the epistemic status of one's beliefs (Hazlett, 2012), a genuine modesty and lack of overconfidence in intellectual strengths (Driver, 2001), and low vanity and low arrogance (Roberts & Wood, 2007).

A potential weakness of the limitations-owning account is that the definition does not include anything that prevents an intellectually humble person from also being arrogant (Church, 2017) and instead argues that a lack of improper pride is only expected to accompany limitations-owning for extra-definitional, psychological reasons (Whitcomb et al., 2017, pp. 20-25). A later published attempt at synthesizing prior work including the limitations-owning account avoided this problem by defining intellectual humility as consisting of two aspects – modesty and accuracy (Tanesini, 2018). Whereas the limitations-owning account has been more

directly influential on psychological measures, it is this broader synthesis by Tanesini that captures the breadth of conceptualizations in psychology, as will be shown.

Though not generally the primary focus of philosophical accounts, philosophers have included psychological descriptions of intellectual humility in their work with greater specificity than most psychological accounts, which tend to rely on a general trait designation (cf. Leary et al, 2017). Accounts of intellectual humility as proper belief confidence and proper higher-order epistemic attitudes situate intellectual humility as cognitive and meta-cognitive in nature (Barrett & Church, 2017; Hazlett, 2012). On the limitations-owning account, intellectual humility results from personal dispositions that span affective, behavioral, and cognitive domains, though this is stated generally without attempting to provide specifics (Whitcomb et al., 2017). Arguably the most complete psychological account of intellectual humility across either psychological or philosophical literature is Tanesini's (2018)—defining intellectual humility as a cluster of attitudes, dedicating a significant portion her paper to a technical description of attitudes as understood in social psychology, and arguing that an attitude cluster is capable of being a virtue.

In psychological research, conceptualizations of intellectual humility reflect a synthesis of prior research on general humility as a personality trait and philosophical accounts of intellectual humility as an intellectual virtue. As with other humility-related constructs, an intellectually humble person is expected to display a humble character (e.g., McElroy et al., 2014). The influence of virtue epistemic accounts is apparent, however, in that expectations of intellectual humility generally involve cognitive sophistication (e.g., considering evidence, recognizing fallibility, thinking carefully, avoiding specific biases; Haggard et al., 2018; Hoyle et al., 2016; Krumrei-Mancuso & Rouse, 2016; Leary et al., 2017; McElroy et al., 2014). Whereas psychological conceptualizations of intellectual humility all describe both humble character and

cognitive sophistication as qualities that will be associated with intellectual humility, there is significant diversity of definitions in the extent to which humble character and cognitive sophistication are core to the trait. At one end of the spectrum, IH has been defined as a subdomain of general humility like any other — possessing a humble character with respect to intellectual matters — without any added cognitive or epistemic elements (McElroy et al., 2014). This conceptualization is aligned with Tanesini's (2018) modesty aspect. At the other end of the continuum, IH has been characterized entirely as cognitive disposition without any reference to emotion or interpersonal behavior (Hoyle et al., 2016; Leary et al., 2017), and this aligns with Tanesini's (2018) accuracy aspect. Most conceptualizations, however, define or characterize intellectual humility as essentially involving both modesty and accuracy aspects — a humble personal character along with cognitive sophistication (Alfano et al., 2017; Haggard et al., 2018; Hill et al., 2019; Krumrei-Mancuso & Rouse, 2016; Zachry et al., 2018). Altogether, then, intellectual humility as a subject of psychological research can generally be thought of as encompassing humble character in intellectual matters (modesty) as well as a cognitive disposition to be aware of and appropriately responsible for one's own epistemic limitations (accuracy).

Challenges for Self-Report Measurement of Intellectual Humility

The question of how intellectual humility might be measured has been addressed from multiple perspectives and has also been a matter of some debate (Church & Samuelson, 2017, pp. 76-78; Church, 2018; Davis et al., 2019; Kotzee, 2015). In discussions of measurement validity, some scholars have noted the possibility of significant bias being introduced through the use of self-report questionnaires because intellectual humility involves accurate self-awareness of limitations and the lack of intellectual humility involves a lack of awareness of limitations

(e.g., Christen et al., 2019; Brienza et al., 2018). This echoes well-known concerns about the seeming paradox of self-attributions of humility and related qualities (e.g., Garcia, 2006; Tangney, 2000). Discussions of this problem in the context of intellectual humility measure development have generally argued that such concerns are addressable through testing and potentially controlling for socially desirable responding (e.g., Church & Samuelson, 2017, pp. 82-83) and by arguing that prior work on general humility has not shown evidence of a modesty effect (e.g., Leary et al., 2017; cf. Zettler, et al., 2016). However, though desirability bias and modesty effects are relevant concerns for humility-related constructs, the far greater challenge is that intellectual humility seems to be required in order to provide an accurate self-report of intellectual humility, even in the absence of any common form of response bias.

The contradiction involved in the self-report assessment of intellectual humility is in relying on people who are unaware of their intellectual limitations to report accurately about the degree to which they are aware of and responsible for their intellectual limitations. Consider the example of a person who dismisses evidence that conflicts with important beliefs as false. When that person is asked how they respond to evidence that challenges an important belief, they will not consider those instances at all because they perceived themselves to be rejecting a falsehood rather than evidence. Such a person might be free from any inclination toward desirable responding yet would nevertheless produce an inaccurate and positively biased self-report. A similar effect of not being able to assess one's limitations has been shown with respect to a variety of skills relevant to intellectual humility (Kruger & Dunning, 1999) including in the assessment of one's own reasoning abilities, which is particularly relevant for how someone might experience themselves in disagreements (Pennycook et al., 2017). That this is common enough to be relevant for measurement considerations is also evident in the consistency with

which highly implausible ideas (e.g., flat Earth theory) are argued by advocates to be readily apparent to anyone who will exercise critical thinking and carefully consider the evidence. In addition to the possibility of excessively favorable self-report, the opposite problem might also occur. Those who are most aware of their limitations will best be able to identify those limitations as they manifest and then recall them when providing a self-report. An intellectually humble person could therefore earnestly rate themselves with lower scores than others who strongly value those qualities, and this would be due to superior accuracy rather than any attempt to downplay strengths. These potential complications draw attention to the need to carefully consider the response process involved in the direct self-report of intellectual humility.

In self-report measures of intellectual humility, items focus on behaviors that are generally expected of any reasonable person — considering opposing opinions (Porter & Schumann, 2018), revising beliefs to align with evidence (Leary et al, 2017), acknowledging fallibility (Hoyle et al., 2016), being respectful of others during a disagreement (Krumrei-Mancuso et al., 2016), caring about the truth (Haggard et al., 2018), and so on. People who are strongly invested in being correct — for whatever reason — are likely to try to embody those qualities if for no other reason than to be correct more often. Exhibiting these qualities consistently, however, requires much more than a strong desire to be correct, and the final element required for accurate self-evaluation is an assessment of how consistently these are applied in practice. For this component of the response process, those who are lowest in intellectual humility are likely to be least aware of their limitations with respect to intellectual humility and provide a correspondingly inflated self-report (e.g. “I am always reasonable”). Those who are high in intellectual humility are most likely at this point to reflect on their own fallibility and weigh that in their response (e.g. “I’m prone to bias just like anyone else”).

Altogether this suggests that, among people who strongly value qualities related to intellectual humility, intellectual humility may have an opposite influence on scores than supposed[1].

Limitations of Traditional Validation Practices

If this problem exists, traditional questionnaire validation practices are unlikely to have detected it for several reasons. One complication for detecting this problem in validation studies is that merely recognizing the epistemic value of intellectual humility is likely to be associated with outcomes relevant to intellectual humility, because beliefs about knowledge and knowing are predictive of intellectual strengths and positive epistemic outcomes (Baron, 1991; Sandoval et al., 2016). For example, even someone low in intellectual humility might have strong critical thinking skills that lead them to recognize the value of considering the opinions of others, acknowledging error, and revising opinions when presented with new (Bailin et al., 1999; Toplak et al., 2014). In contrast, someone who lacks intellectual humility due to willful anti-intellectualism would likely not endorse those qualities. In this way, many expected correlates of intellectual humility should also be expected to correlate with merely valuing intellectual humility. Factor analytic evidence is also unlikely to have identified this problem because, when item content fails to adequately target the intended construct as expected, conventional practices of questionnaire development can nevertheless produce encouraging-seeming structural evidence of validity (Maul, 2017; Rhemtulla et al., 2017). Because even serious problems of measurement validity can go unnoticed in the course of normal psychological research if the assumptions underlying a measurement claim are not considered carefully, the accumulation of publications

[1] This idea is referred to differently throughout the manuscript for different emphasis—valuing intellectual humility, associates of valuing intellectual humility, specific associates such as rationality or critical thinking, etc. In every case, this idea is what is meant—that scores primarily reflect valuing qualities associated with the accuracy aspect of intellectual humility (i.e. trying to be a reasonable person) and that, among people who strongly hold those values, intellectual humility may be inversely related to scores.

featuring a measure should also not be taken as evidence of the absence of a problem of this kind (Borsboom, 2006).

Discriminant evidence of validity (also called discriminant validity or, for some applications, incremental validity) refers to a form of evidence often intended to show that scores are not reflecting some other construct or set of constructs instead of the target construct (AERA, APA, & NCME, 2014, pp. 16-17). When addressing whether scores reflect the target construct rather than related or confounding constructs, discriminant evidence frequently takes the form of statistical analyses showing that the scores under investigation predict unique variance in relevant outcomes beyond what is predicted by measures of rival or confounding constructs (Hunsley & Meyer, 2003). This approach is limited, however, in that the statistical methods employed are prone to produce favorable-seeming but spurious results when measurement error is not modeled in the analysis, something often overlooked in psychological research and also not attempted in the intellectual humility literature thus far (Westfall & Yarkoni, 2016). At the same time, modeling measurement error in such analyses can introduce additional and potentially worse problems when the measurement model is inaccurate, which is common, and addressing this problem can involve methodological roadblocks (Rhemtulla et al., 2020). For these reasons, the question of whether validity evidence demonstrated for an intellectual humility measure is attributable to merely valuing intellectual humility is not likely to have been adequately addressed by any single statistical demonstration and instead requires theoretical attention and broad consideration of all evidence.

Evidence Needed to Support Self-Report Measures of Intellectual Humility

In part due to the generally confirmatory nature of initial validation practices (Kane, 2006; Maul, 2017), a relatively recent advancement in the measurement validity literature has

been to emphasize the importance of evaluating validity evidence with a focus on the primary challenge to the proposed interpretation of scores (Kane, 2013). This builds upon prior guidance that the priority for validity testing is addressing the most plausible and high-impact threats to the validity argument (Cronbach, 1989; Crooks, et al., 1996). Having argued here for a specific threat to validity that is both highly likely and high impact, for this review, I consider the most important quality of validity evidence for a self-report measure of intellectual humility to be the extent to which that evidence would not be expected of merely valuing intellectual humility.

The clearest such evidence will be association with behaviors or observations that are clear and direct displays of intellectual humility. Examples of such behaviors noted in formal philosophical accounts of the virtue include admitting limitations to self and others when doing so is otherwise difficult (Roberts & Wood, 2007; Tanesini, 2018; Whitcomb et al., 2017), not being excessively confident in one's own beliefs (Church, 2016; Hazlett, 2012; Roberts & Wood, 2003; Whitcomb et al., 2017), seeking or accepting help when needed on intellectual matters (Whitcomb et al., 2017), appropriately making concessions in intellectual disagreement (Hazlett, 2012; Tanesini, 2018), displaying intellectual modesty and low egotism (Roberts & Wood, 2003; Tanesini, 2018), appropriately qualifying claims according to the supporting evidence (Schwab, 2012), acknowledgement of uncertainty or ignorance (Church & Barrett, 2016; Hazlett, 2012; Whitcomb et al., 2017), and attempting to remedy intellectual limitations (Whitcomb et al., 2017). Though this is not a comprehensive list of behaviors characteristic of intellectual humility, this set of examples provides an idea of what would clearly indicate intellectual humility in contrast with merely valuing intellectual humility.

Because direct displays of intellectual humility such as these can be difficult to facilitate in the context of psychological research with large samples, survey-based studies may rely on an

inference to intellectual humility from associated attributes. Examples of such attributes include high intellectual engagement, knowledge and intelligence, critical thinking ability, appropriate skepticism, enthusiasm about science or rational discourse, open-mindedness, agreeableness, curiosity, proper belief commitment, and myriad forms of cognitive sophistication (media literacy, science understanding, specific expertise, etc.). Because these qualities are generally also expected to be associated with valuing intellectual humility, observations related to these attributes cannot be assumed to reflect intellectual humility without a convincing supporting argument. This argument will be more straightforward to the degree that context makes intellectual humility salient. For example, critical thinking is more relevant to intellectual humility when it is applied to one's own views. Providing a charitable (and empathetic) account of an opposing view is more relevant than simply knowing about an opposing view.

With these considerations in mind, the strongest form of evidence in support of a measure of intellectual humility will be clear association with observations characteristic of intellectual humility that cannot be attributed to qualities such as intelligence, critical thinking abilities, or high engagement. Though much less compelling than the strongest form of evidence, the next most promising form of evidence will involve observations that are associated with intellectual humility but which could alternatively be attributed to qualities such as intelligence, critical thinking abilities, or high engagement. Less clear to interpret will be observations that are clearly attributable to qualities such as intelligence, critical thinking abilities, or high intellectual engagement. Due to difficulties with interpretation with respect to the primary threat to validity, significantly less compelling than direct observations will be generally favorable correlations with relevant self-report trait measures. Least persuasive will be correlations with loosely related trait measures.

It is important to note that these standards of evidence pertain to addressing a particular serious threat to validity. Addressing this threat is a necessary but not sufficient condition of an overall demonstration of validity. In addition to showing that scores are associated with intellectual humility and not merely with valuing intellectual humility, a complete validity argument for a measure of trait intellectual humility would need to establish that scores perform well across the domain of behaviors that characterize intellectual humility and do so across relevant situational contexts. Because such arguments generally require support from a substantial program of research, at this early stage, the priority is on evaluating the promise of evidence gathered so far (Kane, 2013). Furthermore, the identified threat concerns reporting on one's intellectual limitations, so for the two measures that do not involve intellectual limitations, evaluation will focus more on general standards of validity evidence (e.g., Messick, 1989).

Because fundamental problems with self-report have not been adequately addressed thus far in validity testing for intellectual humility measures or in reviews of such measures, I set out to conduct a critical review of intellectual humility measures with a focus on evaluating evidence with these challenges in mind. Additionally, during this research, an analysis of descriptions of the intellectual humility construct across measurement papers was conducted, and the resulting content areas and descriptions are also reported here.

Method

Literature Search and Inclusion Criteria

I conducted a search on August 16, 2021 using PsychInfo for articles and citations containing the phrase *intellectual humility* in the title, abstract, or article text, returning 145 results. Another search was conducted on the same date using ProQuest Social Science Premium Collection for same phrase anywhere in the text and limiting results to scholarly journals,

returning 344 results. The titles and abstracts of these articles were reviewed for references to the empirical study of intellectual humility or like concepts (e.g. willingness to admit to being wrong), and those articles were checked for a questionnaire measure of trait intellectual humility (including self-report and other-report questionnaires but excluding coding schemes, state measures, and situated measures). This resulted in the identification of 14 distinct questionnaire measures and additional variant adaptations of these (e.g., slight modifications to target narrower content domains). These articles, as well as reviews of intellectual humility, were also checked for reference to measures that may have been missed by the search, resulting in the addition of one measure. In total, 15 distinct trait questionnaires were identified through the search. Of these, measures with published validation studies were selected for inclusion, resulting in 10 questionnaire measures of trait intellectual humility to be reviewed.

Reporting Procedure

For each instrument, the published validity argument was reviewed. First, the definition and general theory accompanying the conceptualization and operationalization of the construct are described. Second, the basic structure of each measure is reported, including information regarding the number and interpretation of factors, the number of items in each scale and subscale, as well the item that accounts for the most subscale variance in the reported factor analysis. Third, I report the reported range on the Cronbach's alpha coefficient, as well as temporal stability. If applicable, I also report results of measurement invariance tests. Fourth, I report other evidence of validity presented by measure authors, including convergent, discriminant, and criterion-related evidence.[2]

[2] For the sake of brevity, evidence is discussed as having been displayed for or by the test (e.g. "The measure showed evidence of validity in relationships with…"), but this is more properly expressed as, "the proposed interpretation of scores attached to the measure was supported by scores being related with…"

Results of Review

A summary description and critical review of validity evidence for each scale follows. Content domain descriptions were organized into themes inspired by previous work on general humility measures (McElroy-Heltzel et al., 2019) and are provided in Table 1.1. Summary data are reported in Table 1.2.

General Intellectual Humility Scale (GIHS)

The General Intellectual Humility Scale (GIHS; Leary et al., 2017) is a self-report measure that consists of six items with response choices arranged on a 5-point Likert-like scale ranging from *not at all like me* to *very like me*. The authors of the GIHS define intellectual humility as "recognizing that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in obtaining and evaluating relevant information" (Leary et al., p. 793).

This is a narrower conceptualization than those employed by developers of other measures of general intellectual humility and is unique among such measures in its exclusive focus on epistemic and cognitive aspects of IH. The GIHS conceptualization features cognitive and metacognitive processes (monitoring, recognition, attending, evaluating), epistemic considerations (doubt, belief revision, assigning trust) and identifies IH with favorable epistemic outcomes (recognizing fallible beliefs in light of evidence). This represents one end of the spectrum of conceptualizations in contrast to other measures that all include references to either interpersonal or affective qualities and lack any reference to justified or fallible beliefs instead of, for instance, intellectual fallibility in general.

The GIHS is a single-factor instrument comprised of items such as "I accept that my beliefs and attitudes may be wrong." Cronbach's alpha values for GIHS scores ranged from .73

to .87 across six samples. Temporal stability was not reported given that no longitudinal study was included. In the initial validation studies, Leary et al. (2017) showed convergent evidence of validity in confirming expected correlations between the GIHS and openness (r = .33), the interest component of epistemic curiosity (r = .35), existential quest (r = .35), need for cognition (r = .34), dogmatism (r = -.49), intolerance of ambiguity (r = -.32), and self-righteousness (r = -.35). Other effects were found for self-attributions of intellectual humility (r =.37) and open-mindedness (r =.35). There were two results that did not align with the authors' predictions and that were accordingly noted as surprising. The GIHS failed to show the expected relationships with narcissism (expected a negative relationship, no relationship found) and social vigilantism (expected a negative relationship, no relationship found).

The GIHS showed initial criterion-related evidence of validity in a series of three experimental studies, yet some predicted results were not observed. In the first study, multiple regression was conducted involving GIHS scores, religiosity, experimental condition (reading pro-, anti-, or balanced essays about religion), and response to reading a persuasive essay on religion (affect, agreement, perceived accuracy, impressions of the author). Results showed that the GIHS predicted less extreme religious views, less certainty and superiority about one's own religious views, greater agreement with essay arguments, more favorable evaluation of the accuracy of balanced and anti-religious arguments, preference for balanced rather than one-sided arguments, and more positive character judgements about the authors across essay conditions. The authors noted that, though these findings were generally consistent with expectations of IH, examination of the interaction between the article condition and IH revealed the absence of the expected effect of IH in predicting more favorable evaluations of argument accuracy in the pro-religion essay condition. The authors proposed this could be due to their sample being more

prone to agree with the pro-religion essay (23% of the sample disagreed with the pro-religion essay whereas 56% disagreed with the anti-religion essay), leaving less room for IH to have an effect in moderating the influence of disagreeableness.

In the second study, the GIHS was tested for its ability to predict more favorable attitudes toward people who change their beliefs (Leary et al., 2017). Participants were asked for their political identity affiliation (Republican, Independent, or Democrat) and presented with a fictional political candidate who previously supported an unspecified position about the environment but had now come to hold the opposite view after learning more and realizing he was wrong. Participants were randomly assigned such that the candidate was described as either Republican or Democrat. Results showed that, as expected, those scoring high on the GIHS were more likely to believe and vote for a candidate who had changed his position on an issue and less likely to characterize the change as flip-flopping to win votes. However, the authors noted that this finding was qualified in that the effect was found only for Republicans and that the GIHS did not find the expected effect of a more favorable attitude toward change of beliefs among Democrats or Independents. The authors suggested that the effect observed for Republicans may be the result of especially negative attitudes toward change of political position among Republicans and that IH would therefore be especially relevant to moderating judgements among Republicans in particular. The authors did not offer any additional explanation concerning the absence of the predicted effects of IH among Democrats or Independents.

In the third study, the GIHS was tested in relation to ability to distinguish strong from weak arguments both generally and when the argument is personally relevant to them. Participants were presented with persuasive essays about flossing featuring arguments that were either weak and anecdotal or strong and scientific and were asked to select the most compelling

reasons for flossing from a set of mixed strong and weak reasons. Participants reported how often they floss and in the analysis were categorized into flosser and non-flosser groups.

Results showed that non-flossers high in IH drew clearer distinctions between strong rather than weak arguments in favor of flossing, endorsed the stronger reasons for flossing, and reported being influenced by strong arguments more than by weak arguments. In contrast, non-flossers low in IH did not show these effects. Contrary to predictions, however, these effects were not observed even to a lesser degree among flossers, for whom IH did not predict effectively distinguishing between strong and weak arguments and did not predict reporting being more influenced by strong rather than weak evidence. The authors had predicted that the personal relevance of potentially changing a behavior would motivate flossers to engage the information critically whereas non-flossers would be less motivated due to it being of no practical consequence to their ongoing behavior. Thus, they were correct in that they expected high IH would show more sizable effects when the information was more personally relevant. However, they were incorrect in predicting those high in IH would still show the effects to a lesser degree. Finally, IH did not predict greater reported *intent* to floss in the coming week following reading arguments for why flossing is important. Leary et al. (2017) investigated discriminant validity of the GIHS as well during this third outcome-focused study by comparing the GIHS with need for closure in their analysis. Results provided discriminant evidence of validity in that the two constructs were found to be only weakly related (r = -.14, p = .007) and because the GIHS showed similar relationships with outcomes even when controlling for need for closure.

In addition to specific concerns raised about the results of individual studies, Leary et al. (2017) acknowledged in their discussion that “many of the effect sizes for intellectual humility

were relatively small, which could be interpreted as an indictment of general intellectual humility as a construct" (p. 809). They argued against this conclusion, however, by offering the alternative explanation that effects are small because myriad factors influence "people's reactions to beliefs, attitudes, and people with which they disagree" (Leary et al., 2017, p. 809) and because general intellectual humility is subject to variation in expression across situations and across topics.

Commentary on Evidence for the General Intellectual Humility Scale

The GIHS is supported by a specific account of the nature of intellectual humility as a cognitive disposition. In addition to structural evidence and associations with other trait measures, the validity study features three behavioral studies targeting qualities relevant to if not characteristic of intellectual humility, the most of any validity paper for a measure of intellectual humility. Each of these studies found evidence that could be attributed to intellectual humility, and if the interpretations of that evidence proposed by the authors are warranted, then these findings would arguably constitute a demonstration of GIHS scores being associated with behaviors that are characteristic of intellectual humility as they have conceptualized it. At minimum, then, the GIHS demonstrated in initial validity testing a clear capacity for reflecting qualities associated with valuing intellectual humility. Determining whether the evidence supports interpreting scores as reflecting intellectual humility requires a close look at key observations from each study.

In the first of the behavioral studies, findings generally showed GIHS scores to be associated with lesser certainty about and extremity of one's own religious views and a more favorable assessments of essays that included an anti-religious argument and more favorable attributions about the authors of those essays. The key aspect of this study requiring scrutiny is

the lack of effects with respect to the pro-religion essay. Though Leary et al. (2017) explained this lack of effect as a function of fewer people in the sample finding the pro-religion article disagreeable, an alternative explanation would be to take at face value the result that GIHS scores do not predict greater acceptance of pro-religion articles among people who find religion disagreeable. Whatever the reason for not observing a more favorable evaluation of pro-religion articles, the lack of demonstrating this finding is consistent with GIHS scores primarily reflecting valuing empirical evidence and rationality. This interpretation would be consistent with the more favorable evaluation of anti-religion articles even for religious people — believers who place strong value on evidence and rationality are likely to have had practice engaging with anti-religion opinions and thus have a less negative response to the anti-religion articles. If the GIHS were reflecting the ability to engage with disagreeable information and not just holding values that rational discourse and empirical evidence are good, then scores should specifically predict more moderate evaluation of pro-religion essays that strain anti-religion worldviews, but this was not observed. It may be that the GIHS can demonstrate this property in subsequent research with a diverse sample with respect to religious views.

A similar pattern was observed in that the GIHS was found to predict the expected intellectual humility-relevant outcomes among Republicans but not among Democrats and Independents. Leary et al. (2017) explained this finding by arguing that, because ideological commitments that characterize conservatism result in negative attitudes toward change of beliefs, high GIHS scores among Republicans are noteworthy departures from group norms in a way not true of Independents and Democrats. This argument from Leary et al. favors the alternative interpretation of GIHS scores as primarily reflecting epistemic values (i.e., it is good to change one’s belief in response to evidence and reason) rather than reflecting intellectual humility.

Between this finding and the findings from the study on religious beliefs, it seems that GIHS scores may be more predictive of outcomes relevant to intellectual humility in groups where strongly valuing reason and empirical evidence is not characteristic of the group.

The third study demonstrates a general effect that people use their critical thinking abilities more when presented with information that suggests an individual may need to make changes than when presented with information that is more congenial to existing beliefs and behaviors (Ditto et al., 1998). Demonstrating that this effect is associated with GIHS scores is evidence that scores predict a willingness or ability to engage with arguments and evidence, which may be a necessary component of intellectual humility. However, this falls short as a demonstration of intellectual humility because GIHS scores did not predict a disposition to engage in critical thinking skills to evaluate congenial information, which would have been a more noteworthy result with respect to intellectual humility.

Altogether, these results suggest that the GIHS reflects epistemic values associated with intellectual humility, but the evidence does not clearly support GIHS scores as reflecting intellectual humility in the cases that could differentiate this from intellectual humility. It may be that, for individuals where even holding those values involves a strain against group norms or a tension with other important values, GIHS scores will be more reflective of intellectual humility in that group. In contrast, in groups where valuing intellectual humility is readily compatible with group values, the GIHS may not be very predictive of intellectual humility. Additional research is needed to investigate the properties of GIHS scores in various groups, but the clarity of the conceptualization and clear evidence of results consistent with intellectual humility should make the GIHS a focus of further research.

Specific Intellectual Humility Scale

The Specific Intellectual Humility Scale (SIHS; Hoyle et al., 2016) is presented in both 9-item and 3-item forms with responses on a 5-point scale anchored by *not at all like me* and *very like me*. The SIHS was developed to assess intellectual humility with respect to any given subject matter or topic (Hoyle et al., 2016). Items are designed to have a reference to a specific topic inserted in the text. The SIHS was developed by some of the same researchers as the GIHS, and the definition of intellectual humility on the SIHS is accordingly similar: "the recognition that a particular personal view may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that view and to one's own limitations in obtaining and evaluating information relevant to it" (Hoyle et al., p. 165). As with the GIHS, this definition reflects the same distinctly cognitive, epistemic, and normative conceptualization that contrasts with the more affective and interpersonal definitions attached to instruments authored by other groups.

The SIHS is a single-factor instrument in both short and long form, though Hoyle et al. (2016) describe intellectual humility as involving three component aspects and structured the final scale to include three items per aspect. The brief form of three items is comprised of the strongest-loading item reflecting each aspect: "My views about _____ today may someday turn out to be wrong,"; "When it comes to my views about _____ I may be overlooking evidence."; and "My views about _____ may change with additional evidence or information." Given the goal of functioning as an intellectual humility measure across subject matter domains, tests of measurement invariance across topic were conducted with results supporting the claim of invariant measurement across topics. Cronbach's alpha values reported ranged from .88 to .96 for the full 9-item scale across multiple topics in 14 administrations to two samples and from .77

to .88 for the 3-item version across multiple topics in 20 administrations to two samples. No tests of temporal stability were reported.

Convergent evidence of validity was examined across many topics and by way of multiple tests, and results are reported here in ranges if the range included at least one effect size where $|r| \geq .30$ (Hoyle et al., 2016). The predicted relationship with dogmatism was observed and ranged in effect size from small (r = -.22) to large (r = -.53) across six topics. The correlations between specific intellectual humility and general intellectual humility were small to moderate for five of the six views (r = .24 to r = .43) and large for the sixth view (r = .63) with narrower topics being less strongly related to GIHS than more general topics, suggesting that SIHS is not only merely a reflection of GIHS. In a second study, the effect size for correlations between specific and general intellectual humility across 12 views ranged from small to medium (r = .09 to r = .36) and correlations among SIHS scores were generally more highly correlated between related topics within the same domain than between topics in different domains. Some observed results did not align fully with expectations. The predicted relationship with openness was observed to be only modest (r = .11 to r = .21). Correlations between the SIHS and GIHS were weak in the second study relative to the first study (r = .09 to r = .31), which the authors suggested may be due to the GIHS being completed earlier in the day than the SIHS scales. The second study also failed to find the predicted result that SIHS scores for more granular views would be less strongly related to GIHS than SIHS for more general views.

In criterion-related tests of validity, the SIHS predicted more moderate levels of agreement/disagreement with four issue-specific statements, with SIHS demonstrating a curvilinear relationship such that both strong disagreement and strong agreement corresponded with lower IH whereas moderate responses corresponded with higher levels of IH (Hoyle, 2016).

The SIHS also showed criterion-related evidence of validity in that, as expected, scores predicted some differences in the sources of information people credited for the basis of their viewpoints, with endorsement of "information I learned based on my own exploration/study/research" being related to SIHS for 8 of 12 topics (r = -.33 to r = -.23). Evidence of discriminant validity was demonstrated by comparing the relationship between Dogmatism and SIHS. The authors predicted that SIHS would show small to moderate correlations with Dogmatism, and results showed correlations with effect sizes ranging from r = -.53 to r = -.22.

Commentary on Evidence for the Specific Intellectual Humility Scale.

The SIHS shares similar strengths as the GIHS with respect to the supporting theory and the factor structure. Tests of measurement invariance across topics are also appreciated as part of pursuing a higher standard for psychometric evidence. In addition to these strengths, the most noteworthy feature of the SIHS is the attempt to measure topic-specific intellectual humility rather than intellectual humility as a global trait. This is a promising avenue of research given evidence from epistemic cognition research that there may not be a domain-general cognitive sophistication with respect to the justification and revision of beliefs.

For instruments targeting a global disposition to be open or receptive to evidence, the primary threat to validity is that global evaluations will reflect values more than actual disposition. However, in applying similar items to assessing intellectual humility about specific topics, the primary threat to validity is conflating intellectual humility with low interest and investment in the topic, and it is not clear that results from testing for the SIHS reflect more than the presence or absence of strongly held beliefs on a topic. Because it is generally appropriate for strongly held beliefs to be less open to revision (Oaksford & Chater, 2009), it makes sense to rule out the possibility of low investment in a belief before concluding that willingness to revise

a belief is a sign of intellectual humility, but strength of prior belief was not controlled for in these studies.

Results from the study on information sources cited as the basis for holding specific beliefs further support this alternate interpretation. Reporting that views were partly based on personal investigation (doing research, seeking information) was inversely related with SIHS. This seems strongly counter to expectations of intellectual humility because intellectual humility should be positively associated with seeking information as part of forming a belief about substantive issues. However, such a result is consistent with interpreting high scores on the SIHS as reflecting indifference, ignorance, or low importance of a topic.

Further supporting the importance of this limitation, subsequent research has shown that the SIHS unexpectedly predicted results inconsistent with intellectual humility yet consistent with interpreting the SIHS of reflecting strength of belief (Jankowski et al., 2019). Furthermore, a study investigating the effect of SIHS (religion) scores on comfort with participating in a diverse religious group showed that high SIHS scores were negatively associated with endorsing religion as intrinsically important while also being predictive of greater openness to religious diversity (Zhang et al., 2018). Again, this finding is consistent with interpreting the SIHS as reflecting openness through a lesser investment in the subject rather than as an openness that would be characteristic of intellectual humility. Altogether, it is unclear that SIHS can be interpreted as reflecting intellectual humility about an issue unless there is a strong form of control on belief commitment, strength of belief, belief importance, and other similar considerations.

Comprehensive Intellectual Humility Scale

The Comprehensive Intellectual Humility Scale (CIHS; Krumrei-Mancuso & Rouse, 2016) is a 22-item self-report instrument in which responses follow a 5-point Likert-like scale (1 = *strongly disagree* to 5 = *strongly agree*). For the CIHS, intellectual humility is defined as "a nonthreatening awareness of one's intellectual fallibility" and is further described as involving "openness to revising one's viewpoints, lack of overconfidence about one's knowledge, respect for the viewpoints of others, and lack of threat in the face of intellectual disagreements" (Krumrei-Mancuso & Rouse, p. 210). As suggested by the descriptor of *comprehensive* in the instrument name, the CIHS conceptualization and resulting item content are broad in their content coverage, encompassing interpersonal and intrapersonal qualities and describing social as well as epistemic dispositions, and spanning the modesty and accuracy aspects of intellectual humility.

The CIHS (Krumrei-Mancuso & Rouse, 2016) was designed to be multi-dimensional in conjunction with the authors' conceptualization of IH, and factor analysis additionally supported the existence of four subscales: (a) Independence of Intellect and Ego ("I feel small when others disagree with me on topics that are close to my heart"), (b) Openness to Revising One's Viewpoint ("I am open to revising my important beliefs in the face of new information"), (c) Respect for Others' Viewpoints ("I can respect others, even if I disagree with them in important ways"), and (d) Lack of Intellectual Overconfidence ("For the most part, others have more to learn from me than I have to learn from them"). Confirmatory factor analysis in the original validation paper showed that total scores most strongly reflect Respect for Others' Viewpoints (factor loading = .988), followed by Openness to Revising One's Viewpoint (factor loading =

.794), then by Independence of Intellect and Ego (factor loading = .487), and finally by Lack of Intellectual Overconfidence (factor loading = .372).

Cronbach's alpha values ranged from .82 to .88 across four samples. Scores from the CIHS were predicted to be moderately stable over time, and consistent with that prediction, test-retest correlations were found to be .75 after one month and .70 after three months for the full scale. Krumrei-Mancuso and Rouse (2016) noted that test-retest correlations were below recommended standards for the subscales at both time points and for the full scale at the 3-month time period.

The CIHS showed convergent evidence of validity in correlations with the Openness and Arrogance subscales of the Intellectual Humility Scale (McElory et al., 2014) modified for self-report (r = .52 and r = -.53, respectively) and in showing the predicted positive yet more modest correlation with measures of general humility, including two aspects of dispositional humility: self-correction (r = .42) and accurate self-perspective (r = .30). Additional convergent evidence was found in the observed predicted relationships with open-minded thinking (r = .56) and openness to experience (r = .40).

The CIHS showed criterion-related evidence of validity in predicting open-minded thinking beyond the variance explained by social desirability and measures of IH together and, in a separate analysis, beyond measures of social desirability and general humility together. The CIHS also significantly predicted openness to experience when controlling for social desirability and individualism and predicted open-minded thinking and tolerance when controlling for self-reported desire to engage in cognitive complexity (understand diverse areas of knowledge, synthesize ideas, and engage in logical thinking). The CIHS showed discriminant evidence of validity in the lack observing a significant correlation with self-regard, in showing a small

positive correlation with self-confidence, in showing no relationship with social conformity, and in only a small correlation being observed for social desirability. Additionally, the CIHS was shown to be less strongly predictive of narcissism and psychological entitlement than were measures of general humility, providing further evidence that the measure reflects a quality distinct from general humility and does so in ways expected by theory (Krumrei-Mancuso & Rouse, 2016).

Commentary on Evidence for the Comprehensive Intellectual Humility Scale

The most compelling evidence in support of the CIHS are the observed correlations with actively open-minded thinking (AOT) and that the CIHS demonstrates a relationship with AOT while also showing overlap with measures of humility, suggesting it reflects an attribute that is more interpersonal than AOT alone. However, given that additive effects were not demonstrated across factors, this suggests the CIHS may be reflecting multiple attributes rather than a trait that is connected to both AOT and humility. Furthermore, though the correlation with AOT is certainly in line with what one would hope for from an intellectual humility measure, the evidentiary value of those correlations is qualified by the relative lack of risk in the predictions—items on the more cognitive subscales on the CIHS are highly similar to items on the AOT measure and items on the interpersonal subscales are highly similar to general humility measures. Though the CIHS was shown to predict some outcomes better than a measure of general humility, what remains to be explored is whether the CIHS can account for distinctly IH-relevant outcomes better than the AOT or a subset of AOT questions. Establishing this requires not only more precisely exploring the theoretical and conceptual overlap and departures from IH and constructs like AOT but also considering how to test the CIHS in a way that probes for those differences. Additionally, because AOT measures explicitly do not measure thinking but instead

reflect valuing a type of thinking (Baron, 2017), associations with AOT measure scores are not useful for addressing the central threat to validity of scores reflecting merely valuing intellectual humility.

In subsequent publications, additional evidence pertaining to the qualities of the CIHS has been reported, allowing for further consideration of these issues identified in initial validity testing. The CIHS has been shown to be favorably associated with a variety of other cognitive constructs including general knowledge, reflective cognition, lower overclaiming, need for cognition, curiosity, and intellectual engagement (Krumrei-Mancuso et al., 2019). Once again, however, these results were not compared against other measures likely to produce comparably favorable results, such as AOT. Furthermore, though similar findings were reported elsewhere for cognitive flexibility and intelligence (Zmigrod et al., 2019), the effect was noted to have not been observed for the more identity-related subscales, again raising the question of whether the favorable evidence for the CIHS results from intellectual humility or from the overlap with AOT at the subscale level. Furthermore, the correlations with the CIHS for cognitive flexibility and intelligence were not additive. Whereas either cognitive flexibility or intellectual ability was enough to result in endorsement of the cognitive subscales of the CIHS, both together did not lead to an even higher score. Zmigrod et al. (2019) concluded from this that either intelligence or cognitive flexibility may be adequate to facilitate intellectual humility. However, it is not clear from either philosophical or psychological theory that intelligence alone facilitates intellectual humility nor why it would fail to be additive with cognitive flexibility if it were to contribute to intellectual humility. An alternative explanation based on a more conservative interpretation of results is that either intellectual or cognitive strength can result in self-attribution of epistemic

virtue. Under what conditions and to what degree self-report of intellectual humility corresponds to behavioral indicators of intellectual humility remains an open question

Though research has shown that the CIHS is associated with other cognitive constructs that promote or are associated with knowledge and accurate judgement, it is unclear that the CIHS contributes to that cluster of constructs in ways that are distinctively characteristic of intellectual humility. The candidate explanation of identity-related reservations about admitting fallibility contributing to positive outcomes was not supported by the research reporting subscale-level effects, and it remains plausible that people who are good at some cognitive tasks care about and endorse being good at admitting to fallibility without necessarily being better at doing so in challenging naturalistic contexts relevant to intellectual humility as a virtue. The CIHS seems to be a promising instrument given its breadth of content and the observed favorable relationships with other valuable cognitive constructs, so it is hoped that future research will demonstrate that the CIHS predicts outcomes characteristic of intellectual humility better than other relevant constructs (e.g., cognitive reflection, AOT, critical thinking ability). Until then, it would be prudent to continue to consider that CIHS scores reflect values rather intellectual humility with respect to accuracy-related outcomes.

Intellectual Humility Scale

The Intellectual Humility Scale is an informant-report measure consisting of 16 items with responses arranged on a 5-point, Likert-like scale (1 = *strongly disagree*, 5 = *strongly agree*). McElroy et al. (2014) conceptualized intellectual humility as a subdomain of humility as a general trait: “Whereas humility refers to a variety of domains, intellectual humility (IH) pertains to one’s knowledge or intellectual influence” (McElroy et al., p. 20). Of all the conceptualizations of IH accompanying measures reviewed in this paper, the IHS

conceptualization is distinctive in that it is aligned almost entirely with interpersonal rather than intrapersonal qualities and with social rather than epistemic outcomes. McElroy et al. suggested, for example, that a person's intellectual humility should be most apparent in interpersonally tense situations involving competing ideas, but they did not emphasize notions of epistemic sophistication, in contrast to conceptualizations attached to most other measures with clearly stated expectations for cognitive correlates (e.g., actively open-minded thinking, argument evaluation, knowledge recognition) and/or epistemic outcomes (e.g., belief certainty, belief justification, avoiding errant claims of knowledge).

Cronbach's alpha values for the full IHS scale were .96 and .94 across two studies (McElroy et al., 2014). Temporal stability was not tested. Factor analyses of the IHS suggested and then confirmed the presence of two subscales consistent with the authors' two-element theoretical conceptualization: Intellectual Openness (e.g., "Seeks out alternative viewpoints") and a reverse-coded factor of Intellectual Arrogance (e.g., "Often becomes angry when their ideas are not implemented").

Convergent evidence of validity was demonstrated in that the IHS showed expected relationships with how much the informant trusted the individual they were rating (r = .74), perceived agreeableness (r = .78), perceived openness (r = .54), perceived conscientiousness (r = .58), and perceived neuroticism (r = -.58). The scale also showed criterion-related evidence of validity. In a separate study, participants were recruited who had recently experienced a betrayal by a religious leader, and it was predicted that IHS ratings would differ in association with the participants' spiritual appraisals of the offense and that IHS ratings would additionally be associated with unforgiveness toward the offender. As predicted by theory, IHS scores were found to be related with attitudes toward God (positive attitudes, r = .31; anger r = -.35) and with

forgiveness/unforgiveness dispositions (avoidance, r = -.54; benevolence motivations, r = .41). The forgiveness/ unforgiveness disposition of revenge motivation was observed to be marginally negatively related (r = -.21, p = .054). Contrary to predictions, however, there was no relationship between IHS ratings and viewing the offense as a desecration (McElroy et al, 2014).

The IHS showed discriminant evidence of validity with respect to modesty and drive. Participants were randomly assigned to one of six conditions and asked to think of someone who either most exemplifies or least exemplifies one of three virtuous qualities: intellectual humility, modesty, or drive. These qualities were not defined for participants. Participants in all conditions then provided IHS ratings for their chosen individual. IHS scores were more strongly related to the condition in which participants thought of someone who exemplified IH than to the conditions involving people exemplifying modesty or drive and furthermore clearly distinguished between those individuals thought of as exemplifying IH and those exemplifying drive. McElroy et al. (2014) also included additional criterion-related evidence of validity in that the IHS was strongly predictive of individuals thought to be strong rather than weak in IH.

Commentary on the Evidence for the Intellectual Humility Scale

Though correlations with informant ratings of broad personality constructs were as predicted, some of these relationships were so large as to raise the question of whether the IHS is markedly different in the context of experiencing a religious offense from other favorable character evaluations such as general humility or agreeableness. In the same way, the specific predictions made for perceptions of IH for a religious leader who had betrayed trust would seem just as reasonable as predictions for agreeableness or most any positive character judgement – the more injurious the offense, the less favorably the offender will be viewed.

The prediction involving forgiveness disposition, on the other hand, does seem more obviously relevant to McElroy et al.'s (2014) conceptualization of intellectual humility – it makes sense that someone perceived as more open to other points of view would engender greater forgiveness with all else being equal. However, because there is no comparison provided with how other perceived traits predict forgiveness, it is not clear that the findings in the study substantiate IHS scores as being more predictive of forgiveness than other pro-social traits not specifically associated with openness to other ideas, such as agreeableness.

These concerns notwithstanding, the face validity of items and factor analytic evidence supporting scale construction suggest that the scale would likely be specifically sensitive to cases where a person was generally agreeable but intellectually defensive. What is less clear is if this scale would generally be sensitive to the opposite case of someone who is disagreeable overall yet fair-minded and open about ideas. This is as much a challenge for informant report in general as it is for the IHS in particular (Funder, 2012). Informant report instruments including the IHS will be more interpretable as reflective of trait IH when the informant is positioned to observe the target individual in situations that strain intellectual humility. The less experience the informant has with the target in such situations, the more IHS scores could be expected to converge with more general perceptions of character.

Finally, it is not clear that the setup for the validity studies strain IH as much as might be suggested by the factors involved in a religious offense by a leader. Though McElroy et al. (2014) correctly pointed out that conflict over important values and ideas will strain IH, it is less clear that a religious leader committing an offense primarily represents a conflict of ideas between that leader and others so much as a violation of moral standards. Additionally, though the definition of intellectual humility provided by McElroy et al. describes accurate self-

awareness of intellectual strengths and weaknesses, the scale itself overwhelmingly describes openness to the perspective of others, which one would expect to be related to but not identical with self-awareness of intellectual limitations. Furthermore, social openness about differing ideas is not the same thing as intellectual openness to differing ideas. The IHS is at minimum a compelling instrument for use in situations where an informant has specific experience with the target in situations that uniquely strain the target's intellectual humility yet enough distance from the context of the assessment to maintain objectivity.

The Limitations-Owning Intellectual Humility Scale

The Limitations-Owning Intellectual Humility Scale (LOIHS; Haggard et al., 2018) is a 12-item self-report instrument with responses arranged on a 9-point Likert-like scale (1 = *strongly disagree;* 9 = *strongly agree*). Haggard et al. (2018) anchored their instrument in the limitations-owning account of intellectual humility (Whitcomb et al., 2017). The limitations-owning account of IH provides a rich and detailed conceptual framework for a practical account of intellectual humility as a psychological trait. By explicitly aligning with this framework, the LOIHS is aligning itself with a more conceptually rich and specific account of IH than some predecessors. Haggard et al. concisely summarized the definition of IH in the limitations-owning account as "owning one's intellectual limitations while being appropriately attentive to them" and explained that being IH involves "affective, motivational, behavioral, and cognitive dispositions" that lead to having accurate beliefs about one's own intellectual limitations and also lead to feeling, acting, and being motivated in "certain ways in certain circumstances" (Haggard et al., 2018, p. 185).

In collaboration with other scholars, Haggard et al. (2018) identified three essential components of intellectual humility: (a) owning one's intellectual limitations, (b) love of

learning, and (c) appropriate discomfort with one's own intellectual limits. Factor analyses supported the final scale being reflective of these three theory-informed factors that guided development: Love of Learning ("When I don't understand something, I try hard to figure it out"), Appropriate Discomfort with Limitations ("When I think about the limitations of what I know, I feel uncomfortable"), and Owning Intellectual Limitations ("When someone points out a mistake in my thinking, I am quick to admit that I was wrong"). Cronbach's alpha values of .86 and .87 were reported in two samples. The test-retest correlation for the full-scale score was .75 after five months.

The LOIHS showed convergent evidence of validity in its relationship with other measures of intellectual humility, including the CIHS (r = .52) and the GIHS (r = .43) and also in correlations with agreeableness (r = .46), conscientiousness (r = .49), and neuroticism (r = -.48). Additional convergent evidence of validity was obtained by way of secondary factor analysis of all included measures of intellectual humility that showed them loading on a single factor (though goodness of fit statistics for the model were not reported). Criterion-related evidence was found in strong relationships with closed-mindedness (r = -.48), assertiveness (r = .43), and authentic pride (r = .41), and smaller effects consistent with theory were observed for dogmatism (r = -.23) and hubristic pride (-.26). However, some results differed from expectations in ways that were noted as potentially concerning. First, one quality highlighted as a likely associate of intellectual humility in the Whitcomb et al. (2017) paper is reflective thinking, yet the LOIHS was unrelated to a measure of cognitive reflection. Second, a surprisingly large positive correlation was observed between LOIHS scores and the Self-Deceptive Enhancement subscale of the Balanced Inventory of Desirable Responding (r = .49; BIDR; Paulhus, 1988).

Discriminant evidence of validity was found in the lack of observed relationships between LOIHS scores and general religiousness, social vigilantism, participant age, and participant sex.

Commentary on the Evidence for the Limitations Owning Intellectual Humility Scale

The primary strength of the LOIHS is its full commitment to the limitations-owning account from Whitcomb et al. (2017), which includes a clear account for how intellectual humility is expected to be related to other qualities and which of those qualities are more central to the trait and which are more distally related. Alignment between LOIHS scores and the Whitcomb et al. (2017) paper thus provides numerous specific, testable claims for whether the measure is functioning as would be expected.

Evidence supporting the LOIHS consists of correlations with a variety of self-report trait measures and with one form of direct observational evidence. The direct observational evidence is the association with scores from the cognitive reflection task, which is clearly related to the limitations-owning account of intellectual humility. Though two of the three LOIHS subscales showed positive correlations with cognitive reflection, these correlations were small and were not apparent at the level of the full-scale score. With respect to the associations with other trait measures, the sizable positive correlation with self-deceptive enhancement suggests that LOIHS scores may be susceptible to reporting overly positive accounts of themselves. If so, the subset of people endorsing LOIHS items without warrant may be the reason for the weak association with cognitive reflection.

In total, the LOIHS demonstrates similar associations with other traits as seen across measures of intellectual humility that include the epistemic aspects of its conceptualization. Inclusion of a test of cognitive reflection provided an opportunity for a relatively clear demonstration of evidence consistent with the proposed interpretation of scores, but the results

from that test were not clearly favorable. Similarly, the selection of a more sophisticated measure of desirable responding is a commendable choice for obtaining clearer validity evidence. Results from that test, however, raise concerns about the influence of general self-enhancement bias on scores, in addition to potential problems with awareness of limitations that are the subject of this review. Additional evidence is generally needed for how LOIHS scores can be interpreted.

The Cultural Humility Scale

The Cultural Humility Scale (CHS; Hook et al., 2013) is an instrument designed and named in conjunction with a different humility-related construct—cultural humility, yet it has subsequently been modified for use in multiple studies as a measure of intellectual humility (Hook et al., 2015). Because it is used as a measure of intellectual humility and has a published validation study, it was included for this review. Even in its unmodified form, the CHS targets a conceptualization of cultural humility that closely aligns with notions of intellectual humility seen in some of the other reviewed measures here. Hook et al. describe cultural humility as seen in the extent to which someone can "strive to be effective but also cultivate a growing awareness that they are inevitably limited in their knowledge and understanding of a client's cultural background" (2013, p. 2). Though the original form of the CHS is supported by validity studies and a validity argument, the modified version has not been separately evaluated for its validity as a measure of intellectual humility (Hook et al. 2015). The validity evidence reported here therefore pertains to the original, unmodified scale.

The CHS demonstrated criterion-related evidence of validity by predicting more favorable perceptions of the working relationship between the respondent and their psychotherapist, the argument being that awareness of having limited knowledge and understanding of other cultures would result in more positive therapy relationships (Hook et al.

2013). In a hierarchical regression analysis, this was observed in the case of retrospective and contemporaneous reports about the therapy relationship in different populations, and the variance accounted for remained significant after controlling for the perceived multicultural competence of the therapist. The CHS was not tested for convergent evidence or discriminant evidence of validity.

Commentary on the Cultural Humility Scale

Though the CHS as a measure of intellectual humility is not supported by a formal validity argument, it has seen repeated use in published research as a flexible stem for measuring various subdomains of humility, use that is largely driven by the research activity of its authors. Given that multiple measures have been developed that aim to specifically measure intellectual humility and which were expressly validated for that purpose, the CHS would seem most clearly suited to assessing intellectual humility about cultural matters, its originally intended purpose. That said, a formal validity argument supporting the use of the item stems and a set of rules for adapting the stems as a universally flexible humility subdomain measure could establish the instrument as a strong candidate for assessment of the modesty aspect of intellectual humility, both generally and with respect to specific subject matter.

Porter and Schumann Intellectual Humility Scale

The Porter and Schumann Intellectual Humility Scale (PSIHS; Porter & Schumann, 2018) is a 9-item measure with response choices on a 7-point scale ranging from *strongly disagree* to *strongly agree*. The authors of the PSIHS define intellectual humility as "a willingness to recognize the limits of one's knowledge and appreciate others' intellectual strengths" (Porter & Schumann, 2018, p. 140). This is an original conceptualization that bridges both the modesty and accuracy aspects of intellectual humility. The PSIHS is comprised of six

positively worded (“I am willing to admit it if I don’t know something”) and three negatively worded (“I feel uncomfortable when someone points out one of my intellectual shortcomings”) items reflecting a single factor when accounting for negatively and positively worded items (Porter & Schumann, 2018, p. 143). Cronbach’s alpha values across four studies were reported as an average of .74. No tests of temporal stability were reported.

The PSIHS showed convergent evidence of validity in relationships with general humility (r = .42), agreeableness (r = .35 to r = .41), growth mindset (r = .23 to r = .42), openness to experience (r = .40), need for cognition (r = .40), modesty (r = .31), learning goals (r = .44) and conscientiousness (r = .30). Scores were also positively correlated, as expected, with need for cognition, openness to experience, and epistemic curiosity, though effect sizes were small. Contrary to expectations, PSIHS scores were not significantly associated with need for cognitive closure or with narcissism. In criterion-related tests of validity, PSIHS scores were associated with attributing respectful intent in a scenario describing a classroom disagreement (r = .40) and reporting greater openness to a dissenter in a scenario involving a disagreement outside of class (r = .48). A second study demonstrated similar results when the scenario involved a discussion with someone holding the opposite view about an important socio-political issue, and PSIHS scores were again associated with respectful attributions about the disagreement (r = .34) and greater openness to learning about the opposing perspective (r = .33). In a third study, PSIHS scores were associated with a greater preference for spending time reading about an opposing view rather than a congenial view (r = .16 to r = .29). Across these studies, PSIHS scores remained a significant predictor of these outcomes when controlling for relevant other study variables. The PSIHS showed discriminant evidence of validity in no association with self-esteem, confidence in intelligence, or gender. Though age was significantly correlated with

PSIHS scores in two of the four studies, controlling for age in analyses did not alter results. Socially desirable responding was also not associated with PSIHS scores, though this was tested only in the preliminary evaluation of the instrument's factor structure and not in the studies reported in the paper.

Commentary on Evidence for the Porter and Schumann Intellectual Humility Scale.

The primary strength of the PSIHS is that it is supported by multiple behavioral observations along a core theme that is clearly related to intellectual humility — openness to an opposing view. The most direct supporting evidence of this claim is the finding that PSIHS scores were associated with more time spent reading about information pertaining to an opposing view. This finding was supported by the observations of scores being associated with more favorable attributions about a disagreement and greater self-reported openness to the opposing view in response to two different presented scenarios. As to the strength of this evidence with respect to the core threat to validity, these observations could be attributed to intellectual humility but could also be attributed to qualities associated with merely valuing intellectual humility. Whether this behavior among high scorers is taken as evidence of intellectual humility depends upon the motivation for engaging the opposing information and the manner in which that information was engaged.

While the authors argue against the idea that those scoring highly on their measure are surveying opposing views to derogate them, an alternate possibility is that reading about opposing views is simply more interesting than reading information that is already familiar for someone who is curious and motivated to learn. Scores were also unassociated with choosing not to engage in the reading task at all, which contrasts somewhat with the expectation of willingness to engage an opposing view included in the definition attached to the measure. The finding that

attitude strength and issue knowledge remained unchanged after reading opposing views is not necessary evidence against intellectual humility being responsible, but those findings are inconsistent with the kind of engagement that would clearly constitute a display of intellectual humility—considering seriously the possibility of being at least partly wrong, engaging with an opposed opinion to become more informed, and revising beliefs or belief confidence accordingly—even when doing so would strain intellectual humility (e.g., during a heated argument).

Altogether, the PSIHS has shown evidence that is consistent with intellectual humility but which is also attributable to valuing intellectual humility. Similar studies including other indicators of the way in which information is engaged and a context making engaging opposing information meaningfully difficult could provide clearer evidence of intellectual humility. Even then, it is worth noting that none of the philosophical accounts of intellectual humility have aligned completely with even virtuous open-mindedness, and open-mindedness alone can even be a sign of a lack of appropriate confidence in beliefs. For the PSIHS to function as a measure of intellectual humility, in addition to clearer evidence that scores predict a virtuous form of open-mindedness characteristic of intellectual humility, a stronger argument for the centrality of open-mindedness to intellectual humility would need to be made or additional evidence demonstrating broader domain coverage would need to be provided.

Alfano Multi-Dimensional Intellectual Humility Scale

The Alfano Multi-Dimensional Intellectual Humility Scale (AMIHS; Alfano et al., 2017) is a 22-item measure with response choices on a 7-point scale ranging from *strongly disagree* to *strongly agree*. Rather than providing a definition of IH, the authors of the AMIHS surveyed several definitions from philosophy and psychology with the intent of developing an instrument

to broadly cover content across the range of conceptualizations. They summarize the range of conceptualizations by describing IH as "a multi-faceted disposition that directs cognition, emotion, and behavior both in social contexts and in solitary inquiry" (Alfano et al., 2017, pp. 2-5).

The AMIHS was designed to be multi-dimensional to reflect theory, and the final scale is comprised of four factors supported by CFA: (a) Open-Mindedness (If I do not know much about some topic, I don't mind being taught about it, even if I know about other topics), (b) Intellectual Modesty (I like to be the smartest person in the room [reversed]), (c) Corrigibility (If someone points out an intellectual mistake that I've made, I tend to get angry [reversed]), and (d) Engagement (I find it boring to discuss things I don't already understand [reversed]; Alfano et al, 2017). In lieu of providing Cronbach's alpha values, the authors presented the results of analyses in the item response theory framework. No tests of temporal stability were reported.

The AMIHS showed convergent evidence of validity in associations between self- and informant-report for each subscale. The strongest correlation between self- and informant-report was for Intellectual Modesty (r = .47). The Open-Mindedness (r = .34), Corrigibility (r = .29) and Engagement (r = .28) subscales showed lesser agreement between self and informant, which the authors noted as lower than typical for personality measures yet to be expected for IH given the largely internal nature of the construct. Additional convergent evidence of validity was demonstrated at the subscale level. As predicted, the Open-Mindedness subscale was negatively related to narcissism (r = -.09) and overclaiming bias (r = -.16), and the Modesty subscale was negatively related to narcissism (r = -.40) and overclaiming bias (r = -.09). The Engagement subscale was positively associated with grit, as predicted (r = .25). Associations with self-

deceptive enhancement and impression management were noted as larger than hoped, and AMIHS subscales correlated more broadly with other constructs than expected.

In a test of criterion-related validity, the Open-Mindedness subscale was tested for its ability to predict overclaiming bias above and beyond personality traits. In a full regression model including Open-Mindedness and all six Big Six personality traits, open-mindedness was most predictive of overclaiming bias. The AMIHS showed discriminant evidence of validity in not showing too large of an effect in bivariate relationships with other study variables and in demonstrating unique predictive value when included in regression analyses with other study variables.

Commentary on Evidence for the Alfano Multi-Dimensional Intellectual Humility Scale.

The AMIHS is supported by one of the most thorough considerations of measurement validity accompanying any measure of intellectual humility — a substantial contribution to theory, testing the instrument with multiple strong forms of evidence, specific predictions for observations, and open acknowledgment of failed predictions. The authors make a compelling argument that intellectual humility as a psychological construct consists of at least four dimensions by reviewing major philosophical accounts for psychological characteristics. This approach builds upon the conceptual rigor of specific philosophical accounts, attends to the breadth of the construct as discussed in psychology, and identifies psychological factors that could plausibly promote a disposition aligned with the broad expectations of intellectual humility. In addition to important theoretical work, the authors also advance the measurement of intellectual humility in addressing whether the items function well at different levels of the target construct, which seems especially important in the study of a virtue.

With respect to the central threat to validity of how scores are to be interpreted, supporting evidence consists of informant-report, correlations with other self-report trait measures, and direct observation of qualities associated with intellectual humility. Informants generally had long-term familiarity with the participant but there were only one or two informants for all but one participant, in contrast to some group-based studies of perceptions of intellectual humility that have been conducted with other measures. The modest correlations between informant report and self-report for most dimensions are not concerning for the reasons given by the authors — the processes described are mostly internal and there may be different interpretations of items as well — but these reasons also preclude global informant report from being strong evidence of validity for any factor other than the Modesty factor. With respect to the modest agreement between self and informant report on the more internal factors, one would expect valuing intellectual humility to be the predominant factor contributing to global evaluations from others, with qualities such as anti-intellectualism, dogmatism, or self-avowed stubbornness being most likely to show clearly in both self-report and informant report. Additionally, informant reports were on average less favorable than self-reports for all dimensions, and this difference was statistically significant for the Open-Mindedness and Engagement subscales.

The direct observational evidence provided is a negative correlation between self-reported open-mindedness and overclaiming bias. It is not obvious that merely valuing intellectual humility would protect against overly confident meta-cognitive assessments about knowledge or to predict a lesser likelihood of over-representing one's knowledge. With respect to its relevance to the central threat to validity, this is a strong form of evidence in favor of the proposed interpretation of scores for the Open-Mindedness subscale. However, the simultaneous

finding that scores from the Open-Mindedness subscale were positively correlated with self-deceptive enhancement to an even greater degree highlights the difficulty of interpreting evidence of one dimension of intellectual humility as evidence of all dimensions of intellectual humility. One of the most interesting features of intellectual humility as a construct in contrast to other forms of cognitive sophistication is that intellectual humility involves a tension between remarkable cognitive strengths and the absence of arrogance or egotism about those strengths. Ideally a measure of intellectual humility will be able to show these qualities simultaneously.

Though correlation with other trait measures is not generally a strong form of evidence, the authors provided specific theory-informed predictions, allowing for the possibility of a relatively stronger demonstration of validity evidence should results match those predictions. Significant departures from predictions, however, resulted in evidence that is not especially informative for addressing the core threat to validity. Correlational evidence was mixed in that it included broad correlations with personality and two forms of desirable responding. The possibility of significant self-report bias seems plausible given the total picture of evidence. In addition to these mixed findings, there are important gaps in criterion-related evidence showing that the instrument is predictive of distinctively IH outcomes or exemplars. Given the many conceptual strengths of this instrument and a promising finding with respect to overclaiming, the AMIHS likely warrants further study to determine its potential as a measure of intellectual humility.

Theistic Intellectual Humility Scale

The Theistic Intellectual Humility Scale (TIHS; Hill et al., 2021) is an 11-item scale with response choices on a 6-point scale ranging from *strongly disagree* to *strongly agree*. The TIHS was developed to assess intellectual humility as it might be understood and experienced by

Christians. The TIHS is conceptually rooted in the philosophy of Augustine, with theistic IH being characterized as "a glad intellectual dependence on God" (Hill et al., 2018, p. 198). This conceptualization of IH differs from others in that it is grounded in a particular religious tradition and is intended for use with members of that faith. The conceptualization shares general features with other IH conceptualizations in that it encompasses both cognitive and intra-/interpersonal content, though these conceptualizations are markedly different in specific content due to being couched within a "Christian grammar" (Hill et al., 2018).

The TIHS is a three-factor instrument: (a) Intellectual Submission to the Divine ("I try to submit all my intellectual efforts to God"), (b) Human Finite Limitations ("I don't need to know everything because God is in control"), and (c) Belief Bias and Limitations ("There are some aspects of my Christian beliefs that I am very confident in, and others that I'm less confident in"; Hill et al., 2021, p. 157). Cronbach's alpha values reported ranged from .83 to .85 across three samples. No tests of temporal stability were reported.

The TIHS showed convergent evidence of validity in full-score and subscale-level associations with a number of conceptually related measures. Human Finite Limitations was shown to be positively related to conscientiousness (r = .33) and negatively related to hubristic pride (r = -.33). Belief Bias and Limitations was positively associated with need for cognition (r = .31). In a test for criterion-related evidence, the TIHS was shown to moderate the relationship between religious commitment and four of five indicators of well-being such that religious commitment had a more favorable influence on well-being when TIHS scores were higher. A positive influence of religious commitment depended upon or was potentiated by higher TIHS scores for flourishing, ultimate meaning, depressive symptoms, and anxiety symptoms. In every

case, this moderation effect was due to either the Intellectual Submission to the Divine subscale or the Human Finite Limitations subscale and not the Belief Bias Limitations subscale.

Discriminant evidence of validity was shown in small to moderate correlations between full scores and subscale-level scores with other IH measures. Additionally, TIHS full scale scores were shown to be unrelated to extraversion and unrelated to social vigilantism. Though subscales were also expected to show the same lack of association, there was a small negative relationship shown between Belief Bias and Limitations and extraversion (r = -.12) and a small positive association between social vigilantism and Intellectual Submission to the Divine (r = .17).

Commentary on Evidence for the Theistic Intellectual Humility Scale.

The TIHS is unique in aligning with a religious and historical account of intellectual humility rather than a contemporary philosophical account. In doing so, the TIHS places debates about the relationship of religious belief to intellectual humility at the center of the conceptualization and highlights important theoretical considerations for measurement of intellectual humility in general. How does culture affect intellectual humility and its measurement? How are accuracy accounts of intellectual humility to account for religious belief? On the subdomain account of intellectual humility (aligned with the modesty aspect), it would seem that lesser self-focus and low concern for correctness might be compatible with a diversity of ideas about knowledge depending on cultural context and cultural ideas about the nature of truth. Consideration of diversity in ideas about knowledge has relevance far beyond religion, however, including openness to other ways of knowing, and instruments that align solely with strongly empirical ideas about knowledge and the accuracy aspect may be challenged not only by religious and spiritual accounts of truth but by more aesthetic and emotional accounts as well.

The validity evidence provided for the TIHS takes the form of associations with other trait measures, and overall evidence supporting the accuracy-related subscale is weak as a result. However, it is not clear that the accuracy-related aspect of the measure is central to the goals of how the instrument should be interpreted and used. If the primary purpose is to assess intellectual humility within a Christian grammar of submission to God in intellectual matters, secular conceptualizations of owning epistemic limitations may not be as relevant to the crux of the construct, especially with for the subjective wellbeing criterion focused on by Hill et al. (2021). Consistent with this idea, evidence for the TIHS suggests that scores at minimum reflect a sense of comfort and coherence about intellectual limitations from a particular Christian tradition, and that this comfort is related to subjective reports of wellbeing. The lack of involvement from the factor of the scale that reflects the secular notions of owning limitations, however, indicates that theistic intellectual humility may not entail comfort with intellectual limitations in the general sense expected by other accounts. If TIHS scores are intended to reflect comfort with a distinctly Christian notion of intellectual humility, evidence supports it as being likely useful for that purpose. To fully support the proposed interpretation of scores reflecting glad intellectual dependence on God, however, this initial self-report evidence would need to be supported by other forms of evidence including informant-report and relevant behavioral evidence. A particular limitation of the current evidence is that the broad spectrum of positive correlations with desirable traits suggests that TIHS scores may be subject to social desirability effects, even with respect to self-reported wellbeing.

If scores are also expected to function as a culturally responsive measure of intellectual humility, additional evidence is required to demonstrate that the general expectations of intellectual humility involving humble character and accurate awareness of intellectual

limitations are reflected by scores. Studies attempting to link TIHS scores with evidence of intellectual humility as conceptualized by contemporary philosophers will need to take into account cultural considerations. Items that ask Christians directly about willingness to change beliefs, for example, will likely run into identity-related confounds (Stanovich & Toplak, 2019). Using science-related measures as evidence of intellectual humility will also need to be careful to consider cultural elements (Kahan, 2015). Informant-report or coders judging behavior will need to be familiar with cultural considerations along these lines because some statements that would seem to lack intellectual humility (e.g., saying that no evidence can ever change a particular belief) might be compatible with intellectual humility due to contextual factors, as has been argued for some other examples of refusing to revise a belief for personal reasons (Church & Samuelson, 2017, pp. 21-22). It is also important to note that these cultural considerations are applicable to any measure of intellectual humility when used with populations that include religious people.

State-Trait Intellectual Humility Scale

The State-Trait Intellectual Humility Scale (STIHS; Zachry et al., 2018) is an 11-item measure with response choices on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The instrument was developed to assess "core features of IH at both the state and trait level and in daily life" using the same item set and to achieve broad content coverage (Zachry et al., 2018, p. 1408). The authors of the STIHS align their instrument with the limitations-owning account of Whitcomb et al. (2017) while expanding item content to capture aspects of daily life, and they characterize intellectual humility as "a proper awareness of and attentiveness to intellectual limitations" (Zachry et al., 2018, p. 1408).

The STIHS is a unidimensional measure with item content describing a positive and active relationship with challenging one's ideas or having one's ideas challenged by others (e.g. "I search actively for reasons why my beliefs might be wrong"[3]). Cronbach's alpha for the final version of the trait measure was reported as .91. No tests of temporal stability were reported.

The trait version of the STIHS showed convergent evidence of validity in relationships with a broad range of related constructs. In the pilot study with the pilot form of the trait version, these included need for cognition (.47); attributes from the Moral Trait Scale including general moral character (.31), honesty (.31), compassion (.32), and humility (.30); and attributes from the Intellect Scale including overall scores (.66), think (.63), learn (.65), create (.57), seek (.63), and conquor (.62). In the study with the final form of the trait measure, correlates included low dogmatism (-.42); authentic pride (.30); several subscales of the HEXACO personality inventory including Extraversion (.36), Agreeableness (.48), Conscientiousness (.31), Openness (.49), and Altruism (.39); prestige (.33); existential quest (.36); epistemic curiosity (.48); low self-righteousness (-.55); low need for closure (-.30); actively open-minded thinking (.56), and tolerance from the Jackson Personality Inventory (.60).

In a demonstration of criterion-related validity, trait STIHS scores were correlated positively with aggregated state measure scores (.47), though for this study, a coding error caused two items to be excluded from the analysis, potentially altering the scale properties. The state version of the measure was shown through experience sampling to predict more positive evaluations of a self-selected challenging situation involving another person. The state version of the scale was also shown to have a similar factor structure to the trait version, and this was argued to show a link between trait scores and behavior. The STIHS showed discriminant

[3] Item-factor correlations were not reported for this measure. As a result, this item was subjectively chosen as representative rather than selected based on factor loading.

evidence of validity in the lack of association with trait narcissism, emotionality, or religious/spiritual beliefs. STIHS scores were also shown to be significantly predictive of multiple relevant outcomes when entered in hierarchical regression analyses after a competing measure of intellectual humility, and this was demonstrated against GIHS scores and again with CIHS scores.

Commentary on Evidence for the State-Trait Intellectual Humility Scale.

A strength of the STIHS is that there are parallel forms of the instrument to assess both state and trait intellectual humility. This allows for the use of situated experience sampling with the same item set as is used for the trait measure, which provides a unique opportunity for collecting validity evidence. This method has an advantage with respect to problems in the response process compared to other methods in that it prevents someone high in intellectual humility from recalling a relatively rare sort of event as a reason for self-rating with lower scores, as might happen in a global self-evaluation. This paper is therefore a major contribution to the measurement strategies available to intellectual humility researchers if for no other reason than pointing toward a technique to protect against a portion of the threat to the response process, yet for this review, the focus is on validity for the included trait measure and the evidence supporting an interpretation of those scores as reflective of trait intellectual humility.

The primary form of validity evidence for the trait version of the STIHS comes in the form of correlations with other trait measures. To the credit of the study authors, predictions about these correlations were made with specificity and each prediction was provided with a rationale, which substantially increases the potential value of such correlations if predictions are realized with similar specificity. Results were mixed in aligning with predictions, however. The STIHS was shown to correlate with a broad range of constructs including many that are relevant

to intellectual humility and many others only loosely related to intellectual humility. In addition to these correlations, STIHS scores showed sizable correlations with some theoretically irrelevant constructs and small to medium-sized positive correlations with three forms of desirable responding. Altogether, this pattern of correlations does not provide adequate specificity for differentiating between intellectual humility and merely valuing intellectual humility.

The relationship between trait scores and data from experience sampling was also presented as validity evidence, but this evidence is difficult to interpret as validity evidence for the trait measure for a few reasons. First, this form of evidence sits somewhere between direct observation and correlation with another self-report trait measure. Though experience sampling constrains the response process to specific events, it still relies on self-determination of what constitutes a challenging event worth reporting and an accurate characterization of how one responded to the event. Second, the validity argument for the state measure is difficult to interpret. State measure scores were positively associated with favorable impressions of the other person involved in the challenging event, lesser likelihood of perceiving the event as a disagreement, and with more positive evaluations of the nature of the interaction. However, item content for the state measure largely assumes the occurrence of a disagreement as a predicate (e.g., “I reviewed the challenging of my ideas as an opportunity to grow”, “I complimented the ideas of those who disagreed with me”). It is not clear how responses to many of the items should be interpreted during interactions that are not disagreements, and it is also not clear how to interpret some items in the context of a more emotional form of conflict, such as a heated argument with a spouse about a frequent point of conflict (e.g., “I was easily convinced to adopt new attitudes or beliefs”, “I enjoyed trying to make sense of conflicted information”). Finally, it

is not apparent that the similar factor structure of the trait and state versions of the scale establishes a link from self-report of the trait to behavioral evidence of the trait, as it seems is claimed in the paper. A situated self-report is not a behavioral observation, and even if it were, a similar factor structure does not allow for validity evidence pertaining to one latent variable to extend to the other. Demonstrating correlations between trait measure scores and the aggregated scores from the state measure version of the instrument suggests some stability in how people respond to items of this kind, but it is unclear that this shared variance is attributable to intellectual humility.

In total, while the large correlations with intellect, open-minded thinking, and similar constructs suggest that scores indeed reflect valuing intellectual humility, the breadth of other correlations and correlations with desirable responding make it difficult to interpret these results with respect to the central threat to validity. The method of sampling from specific experiences and aggregating those samples is a promising addition to the toolset of intellectual humility researchers wanting to limit sources of error in measurement, but it seems that item content or the instruction set might need to be refined so that the response process is interpretable for the situations in which the state measurements occur. Furthermore, the application of that method in this case is difficult to interpret as validity evidence for how the trait measures scores can be interpreted. Altogether, there is not sufficient evidence to determine that STIHS trait scores reflect intellectual humility rather than merely valuing intellectual humility. However, aggregating state scores could be a promising strategy for investigating intellectual humility in general and as an element of validation studies for trait measures of intellectual humility.

Discussion

Despite being a relatively young area of study, the systematic search revealed the existence of 15 questionnaire measures of trait intellectual humility. After excluding measures lacking a published validation study, 10 measures remained and were included for review. As to the primary focus of the review, validity evidence for each measure was critically evaluated in terms of the definition and theory presented by each author team and with a focus on the primary threat to validity of self-report potentially reflecting merely valuing intellectual humility more than intellectual humility.

One measure, the IHS, differed from all others in both methodology (informant- vs self-report) and in conceptualization (interpersonal vs intrapersonal/cognitive). The IHS thus had the more modest goal of reflecting judgements of other people as generally intellectually open and lacking arrogance. Though the IHS had its own problems of not sufficiently demonstrating that other-appraisals of IH are distinct from appraisals of general humility and general agreeableness, the instrument nevertheless seems suitable for use as an assessment of the more interpersonal conceptualization of intellectual humility as a subdomain of general humility conceptualized as a character judgement. Scores from multiple informants when informants are able to evaluate the individual across multiple opportunities for displaying IH may be able to provide a general judgement of how the person is experienced by others who know them well. Even given the expressly interpersonal conceptualization of IH featured in their definition, McElroy et al. (2014) nevertheless included greater likelihood of holding evidence-based beliefs in their descriptions of IH, and because this was not demonstrated, additional validity is still needed for the IHS if it is to extend beyond interpersonal perceptions to reflecting the disposition of the target being rated.

The remaining nine measures are in the form of a self-report instrument purported to assess intellectual humility as a dispositional trait, with one of those measures having an accompanying state measure as well. One of these measures, the CHS, exclusively focused on the modesty aspect of intellectual humility, and two, the GIHS an SIHS, focused exclusively on the accuracy aspect of intellectual humility. The remaining measures included both the accuracy and modesty aspects in the scale construction. Associations with other self-report measures comprised the largest portion of validity evidence for these measures, and for several instruments, these correlations were the only form of evidence provided to link scores to the behaviors expected of intellectual humility. Association with the aggregate of situated state self-reports of intellectual humility was a distinctive approach among demonstrations of association with other self-report measures. The next most common form of evidence was association with scores representing opinions, attitudes, preferences, attributions about scenarios or people, and similar direct communications. Though less common, a few validity studies included an objectively scored measure interpretable in terms of cognition or meta-cognition, specifically measures of reflective cognition and overclaiming/accuracy. Informant report was gathered as validity evidence for only one self-report measure, and setting aside questionnaire responses as a form of behavior, direct observation of behavior was gathered as evidence for only one measure.

Measure authors employed a variety of methods for addressing the possibility of biased responding. In every case where bias was examined, the process included one or more questionnaire measures of desirable responding. Though measures of desirable responding were often positively associated with the scores of the self-report intellectual humility measure, this was rarely acknowledged as indicative of a potential problem with validity. For one measure, in

addition to measures of desirable responding, the overclaiming technique and comparison with informant report were also used to investigate potential bias.

Evidence for these instruments was reviewed in consideration of two rival interpretations of self-report scores – scores being reflective of intellectual humility and scores being reflective of valuing intellectual humility. It was found that the evidence provided in validity studies for measures of intellectual humility is generally compatible with either of these accounts. However, in addition to these results, most papers also reported findings that are discrepant with expectations of intellectual humility. Though some discrepant findings are to be expected purely by chance, the consistency with which such findings occurred across studies suggests a genuine issue. The most straightforwardly troubling of these discrepant findings is the repeated positive correlation between scores of intellectual humility and measures of desirable responding including self-deceptive enhancement, which is an indicator of inflated self-report and inaccurate self-evaluation (Paulhus, 1988). Scores from intellectual humility measures were also often found to correlate broadly and to a sizable degree with loosely related and unrelated variables, suggesting that intellectual humility measures may be susceptible to other forms of method error. More subtle discrepancies were noted in many cases in the observational studies, but while subtle, many of those discrepancies have been argued here to suggest problems for interpreting scores as reflective of intellectual humility rather than merely valuing intellectual humility.

Of the evidence provided in support of existing measures that include the accuracy aspect of intellectual humility, the most impressive results were those indicating qualities that are clearly associated with that aspect of intellectual humility even as they are also associated with valuing intellectual humility. Such findings were presented in support of the GIHS, PSIHS, and AMIHS in initial validation studies and for the CIHS is subsequent research. Though the effects

were only shown for a portion of the sample, the findings that GIHS scores predicted lesser likelihood of derogating opposing arguments and their authors and predicted more favorable views of changing one's mind come closest to a direct display of intellectual humility as anything associated with intellectual humility scores in a validation study. Because these results with GIHS scores were observed only in the context of groups that are ambivalent about globally subjecting one's beliefs to revision (Republicans and Christians), it may be that strongly valuing intellectual humility functions as a better proxy for intellectual humility in such groups.

In summary, though these measures have not yet been shown to reflect intellectual humility in both the modesty and accuracy aspects, all of the instruments attempting to measure the accuracy aspect seem to at least reflect valuing qualities related to intellectual humility, and some have shown promising though ambiguous findings that suggest the possibility of stronger interpretations in the future. It is likely that valuing intellectual humility can serve as an imperfect but useful proxy for intellectual humility in some cases, such as in groups that do not generally value qualities associated with intellectual humility. However, valuing intellectual humility is less likely to be an effective proxy within groups that generally value qualities associated with intellectual humility (engineers, philosophers, college students, etc.). This is a significant limitation, however, because one of the distinctive features of intellectual humility is that it is important and often in need of nurturing even among people who are otherwise quite sophisticated and accomplished with respect to epistemic goods. This is evident in calls for fostering intellectual humility in professions characterized by learning and success, such as medical doctors (Schei et al., 2019), philosophers (Mizrahi, 2016), physicists (Gibson, 2003), and psychological scientists (Franz, 2021). With respect to the long-term project of researching intellectual humility, any measure of intellectual humility that fails to reflect excessive

confidence in highly intellectual people will not only fail to be useful in studying such populations but will tend to produce an intellectually elitist skew in the basic science of intellectual humility.

Limitations of the Review

As with any systematic review, the search procedure may not have identified all instruments. There may be additional unpublished measures or instruments published in an outlet not indexed. Another limitation of this review is that it applies the most thorough reporting and scrutiny only to the evidence presented in the paper introducing the instrument wherein theory, development, and validity arguments were presented. Though some subsequent findings have been noted in the commentary section where applicable, the original validity studies received more emphasis because the validity argument was the focus of the review. Thus, this paper might be said to review measurement development practices in the intellectual humility literature as much as the instruments themselves. However, because the gathering of validity evidence is an ongoing process, it is possible that a strong instrument paired with an inadequate initial validation effort looks weak in this review yet could be supported by a new validity argument from results demonstrated in subsequent research.

Future Research on the Measurement of Trait Intellectual Humility

If valuing intellectual humility is a good enough proxy of intellectual humility for a given research purpose, the strongest measures will be those with items that address values or beliefs rather than ask for self-attributions of behaviors (because self-attributions may be moderate in intellectually humble people who value intellectual humility highly). If that measurement strategy is pursued, however, it will be important to consider the way in which valuing the accuracy elements of intellectual humility differs from valuing open-minded thinking, because

the actively open-minded thinking construct is a well-researched construct that reflects valuing a style of thinking and overlaps highly with the thinking described by some measures of intellectual humility (Baron, 1991). Are there ideas or attitudes or practices that a person with deep intellectual humility would value that are distinct from actively open-minded thinking and which would be expected to produce different results? Assessing valuing intellectual humility may ideally involve a somewhat different item pool than items intended to reflect intellectual humility directly. If similar themes emerge from expert consideration of that question, existing measures could work well either as they are or with slight modification. Future studies examining differences in the associations between IH-relevant outcomes and self-attribution versus the associations between IH-relevant outcomes and statements of value could also be informative.

If the goal is instead to measure trait intellectual humility directly, however, it seems that additional validity testing is required featuring evidence that can differentiate intellectual humility from merely valuing intellectual humility. In conducting additional validity research, while it should be possible to develop questionnaire studies that are more targeted toward addressing the specific threat described in this paper, it may be easier and more informative to involve other research methods. In particular, response process data and other qualitative data in mixed-methods studies could add the necessary context for making stronger and more definitive interpretations of scores. Additionally, collaboration with teams taking other approaches to the study of intellectual humility could allow for scores to be checked against forms of evidence that would be hard to gather otherwise, such as expensive or time-consuming observational studies or experiments (e.g., Hagá & Olson, 2017; Meagher et al, 2015; Reis et al., 2018). When quantitative study designs are the primary means of investigation, the most relevant variables

that could offer alternative accounts for favorable results, such as actively open-minded thinking to reflect values, should be included and run as competing variables in the analysis rather than as outcome variables, and ideally these comparisons would be made using structural equation modeling to control for measurement error if the psychometric models are compatible with structural equation modeling. It may also be informative to compare intellectual humility measures against composites of interpersonal and cognitive qualities associated with intellectual humility.

Increased variety of participant selection approaches might also be helpful for revealing more about the nature of self-attributions of intellectual humility. For a broader perspective, studies using a nationally representative sample would be a helpful way of checking assumptions of how self-reported intellectual humility relates to demographic variables that have until now been based on findings in limited convenience samples. In contrast, more narrowly focusing on theoretically interesting groups (e.g., epistemologists, intellectual narcissists, Q-Anon followers) could be a clear way of studying intellectual humility self-report with respect to particular group characteristics of interest, and this could extend down to the level of case studies.

Intellectual humility researchers would likely benefit from engaging with scholars in closely related fields that have been working on measurement issues for much longer. In addition to actively open-minded thinking, the broad network of subjects that has been grouped under the banner of epistemic cognition is highly relevant to intellectual humility and has invested a great deal of theoretical and empirical work into measurement strategies (Sinatra, 2016). Epistemic cognition may also be a helpful literature to learn from with respect to the challenges and limitations of domain-generality of the accuracy aspects of trait intellectual humility (Hofer et al., 2006; Muis et al., 2006).

Perhaps most importantly, development of the next generation of intellectual humility measures should work from a modern psychometric approach to test development and validation (e.g., Markus & Borsboom, 2013; Newton & Shaw, 2014). The typical process of developing a survey measure described in the intellectual humility literature (e.g., Church & Samuelson, 2017) reflects conventional practice and meets the standards for publication in most psychology journals, but this conventional paradigm of survey development is prone to under-attending to the centrality of theory in psychological measurement such that problems with validity are likely widespread (Borsboom, 2004). Though it may be hard to accept the idea of problematic measurement practices being widespread, there is a significant gap that has widened over the last two decades as validity theorists have moved forward and practices have largely stayed stagnant (Maul, 2017, pp. 9-16). By grounding future measurement work in a modern framework for test development and validation, measurement claims can be trusted to sit upon solid philosophical and methodological foundations and scholars will have a more accurate awareness of the strengths and limitations of measures of intellectual humility.

In conclusion, intellectual humility is arguably as important a topic for research in our present time as any in all of psychology. However, it is also a difficult topic due to the inherent complexities of the subject matter and the resulting challenges for creation and validation of measurement instruments. Existing measures are important first steps toward measuring intellectual humility, but additional work remains to determine exactly what are the features and limitations of these instruments and under what conditions scores from these measures can be expected to reflect the descriptions of intellectual humility found in the literature.

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (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

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (1999). *Standards for educational and psychological testing*. American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford University Press.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of curriculum studies*, *31*(3), 285-302. https://doi.org/10.1080/002202799183133

Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 169–186). Lawrence Erlbaum Associates, Inc.

Baron, J. (2017). Comment on Kahan and Corbin: Can polarization increase with actively open-minded thinking? Research & Politics. https://doi.org/10.1177/2053168016688122

Bernabé-Valero, G., Iborra-Marmolejo, I., Beneyto-Arrojo, M. J., & Senent-Capuz, N. (2018). The moderating role of intellectual humility in the adoption of ICT: A study across life-span. *Frontiers in Psychology*, *9*, 1-6. https://doi.org/10.3389/fpsyg.2018.02433

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425-440. https://doi.org/10.1007/s11336-006-1447-6

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological review*, *111*(4), 1061-1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brienza, J. P., Kung, F. Y., Santos, H. C., Bobocel, D. R., & Grossmann, I. (2018). Wisdom, bias, and balance: Toward a process-sensitive measurement of wisdom-related cognition. *Journal of Personality and Social Psychology*, *115*(6), 1093. https://doi.org/10.1037/pspp0000171

Christen, M., Alfano, M., & Robinson, B. (2019). A cross-cultural assessment of the semantic dimensions of intellectual humility. *Ai & Society*, *34*(4), 785-801. https://doi.org/10.1007/s00146-017-0791-7

Church, I. M. (2016). The doxastic account of intellectual humility. *Logos & Episteme*, *7*(4), 413-433. https://doi.org/10.5840/logos-episteme20167441

Church, I. M. (2018). Intellectual Humility and Religious Belief. *Journal of Psychology and Theology*, 46(4), 219-242. https://doi.org/10.1177/0091647118807188

Church, I. M., & Barrett, J. L. (2016). Intellectual humility. In E.L. Worthington Jr., D. E. Davis, & J. N. Hook (Eds.), *Handbook of humility* (pp. 78-91). Routledge. https://doi.org/10.4324/9781315660462

Church, I., & Samuelson, P. (2017). *Intellectual humility: An introduction to the philosophy and science*. Bloomsbury Publishing.

Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys* (pp. 147–171). University of Illinois Press.

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, *3*(3), 265-286. https://doi.org/10.1080/0969594960030302

Davis, D. E., Hook, J. N., & Hill, P. C. (2019). Seven challenges of an interdisciplinary project to measure intellectual humility. *Journal of Psychology and Christianity*, *38*(3), 148-156.

Davis, D. E., Worthington Jr, E. L., & Hook, J. N. (2010). Humility: Review of measurement strategies and conceptualization as personality judgment. *The Journal of Positive Psychology*, *5*(4), 243–252. https://doi.org/10.1080/17439761003791672

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*(1), 53. https://dx.doi.org/10.1037/0022-3514.75.1.53

Driver, J. 2001. *Uneasy Virtue*. Cambridge University Press. https://doi.org/10.1017/CBO9780511498770

Franz, D. J. (2021). The role of metacognition and motivated reasoning in the response of psychologists to philosophical criticism. Advance online publication. *Journal of Theoretical and Philosophical Psychology*. https://doi.org/10.1037/teo0000167

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*(3), 177-182. https://doi.org/10.1177%2F0963721412445309

Garcia, J. L. (2006). Being unimpressed with ourselves: Reconceiving humility. *Philosophia*, *34*(4), 417-435. https://doi.org/10.1007/s11406-006-9032-x

Hagá, S., & Olson, K. R. (2017). 'If I only had a little humility, I would be perfect': Children's and adults' perceptions of intellectually arrogant, humble, and diffident people. *The Journal of Positive Psychology*, *12*(1), 87–98. https://doi.org/10.1080/17439760.2016.1167943

Haggard, M., Rowatt, W. C., Leman, J. C., Meagher, B., Moore, C., Fergus, T., Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2018). Finding middle ground between intellectual arrogance and intellectual servility: Development and assessment of the limitations-owning intellectual humility scale. *Personality and Individual Differences*, *124*, 184–193. https://doi.org/10.1016/j.paid.2017.12.014

Hazlett, A. (2012). Higher-Order Epistemic Attitudes and Intellectual Humility. Episteme, 9(3), 205-223. https://doi.org/10.1017/epi.2012.11

Hill, P. C., Dunnington, K., & Hall, M. E. L. (2018). Glad intellectual dependence on God: A theistic account of intellectual humility. *Journal of Psychology and Christianity*, *37*(3), 195-204.

Hill, P. C., Lewis Hall, M. E., Wang, D., & Decker, L. A. (2021). Theistic intellectual humility and well-being: does ideological context matter? *The Journal of Positive Psychology, 16*(2), 155-167. https://doi.org/10.1080/17439760.2019.1689424

Hofer, B. K. (2006). Beliefs about knowledge and knowing: Integrating domain specificity and domain generality: A response to Muis, Bendixen, and Haerle (2006). *Educational Psychology Review*, *18*(1), 67–76. https://doi.org/10.1007/s10648-006-9000-9

Hook, J., Davis, D., Van Tongeren, D., Hill, P., Worthington Jr, E., Farrell, J., & Dieke, P. (2015). Intellectual humility and forgiveness of religious leaders. *The Journal of Positive Psychology*, *10*(6), 499–506. https://doi.org/10.1080/17439760.2015.1004554

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. http://dx.doi.org/10.1016/j.paid.2016.03.043

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological assessment*, *15*(4), 446. https://doi.org/10.1037/1040-3590.15.4.446

Jankowski, P. J., Sandage, S. J., Bell, C. A., Ruffing, E. G., & Adams, C. (2019). Humility, Relational Spirituality, and Well-being among Religious Leaders: A Moderated Mediation Model. *Journal of Religion and Health*, *58*(1), 132–152. https://doi.org/10.1007/s10943-018-0580-8

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, *36*, 1–43. https://doi.org/10.1111/pops.12244

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. Understanding and Addressing the Disinformation Ecosystem. *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.2973067

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. https://doi.org/10.1017/bpp.2016.2

Kane, M. T. (2001). Current concerns in validity theory. *Journal of educational Measurement*, *38*(4), 319-342. https://doi.org/10.1111/j.1745-3984.2001.tb01130.x

Kane M. (2006). Validity. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). American Council on Education and Praeger.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review, 42*(4), 448-457. https://doi.org/10.1080/02796015.2013.12087465

Kotzee, B. (2015). Problems of assessment in educating for intellectual virtue. In J. Baehr (Ed.), *Intellectual virtues and education* (pp. 142-160). Routledge.

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), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Krumrei-Mancuso, E. J., Haggard, M. C., LaBouff, J. P., & Rowatt, W. C. (2019). Links between intellectual humility and acquiring knowledge. *The Journal of Positive Psychology, 15*(2), 1–16. 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

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 and Social Psychology Bulletin*, *43*(6), 793-813. https://doi.org/10.1177/0146167217697695

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 51-69. https://doi.org/10.1080/15366367.2017.1348108

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington Jr, E. L., & 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%2F009164711404200103

McElroy-Heltzel, S. E., Davis, D. E., DeBlaere, C., Worthington Jr, E. L., & Hook, J. N. (2019). Embarrassment of riches in the measurement of humility: A critical review of 22 measures. *The Journal of Positive Psychology*, *14*(3), 393-404. https://doi.org/10.1080/17439760.2018.1460686

Meagher, B. R., Leman, J. C., Bias, J. P., Latendresse, S. J., & Rowatt, W. C. (2015). Contrasting self-report and consensus ratings of intellectual humility and arrogance. *Journal of Research in Personality*, *58*, 35-45. https://doi.org/10.1016/j.jrp.2015.07.002

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). American Counseling on Education/Macmillan.

Mizrahi, M. (2016). Why be an intellectually humble philosopher? *Axiomathes*, *26*(2), 205-218. https://doi.org/10.1007/s10516-015-9284-9

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, *18*(1), 3–54. https://doi.org/10.1007/s10648-006-9003-6

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, *32*(1), 69–84. https://doi.org/10.1017/S0140525X09000284

Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package*, *41*, 79586-79587.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*(6), 1774–1784. https://doi.org/10.3758/s13423-017-1242-7

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177%2F0146167202286008

Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J., Snow, N. E., & Jayawickreme, E. (2021). Clarifying the Content of Intellectual Humility: A Systematic Review and Integrative Framework. *Journal of Personality Assessment*, 1-13. https://doi.org/10.1080/00223891.2021.1975725

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

Reis, H. T., Lee, K. Y., O'Keefe, S. D., & Clark, M. S. (2018). Perceived partner responsiveness promotes intellectual humility. *Journal of Experimental Social Psychology*, *79*, 21–33. https://doi.org/10.1016/j.jesp.2018.05.006

Rhemtulla, M., Borsboom, D., & van Bork, R. (2017). How to measure nothing. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 95-97. https://doi.org/10.1080/15366367.2017.1369785

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30. https://doi.org/10.1037/met0000220

Roberts, R., & Wood, J. (2003). Humility and Epistemic Goods. In M. DePaul & L. Zagzebski (Eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (pp. 257-279). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252732.001.0001

Roberts, R., & Wood, J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283675.001.0001

Samuelson, P. L., Jarvinen, M. J., Paulus, T. B., Church, I. M., Hardy, S. A., & Barrett, J. L. (2015). Implicit theories of intellectual virtues and vices: A focus on intellectual humility. *The Journal of Positive Psychology*, *10*(5), 389-406. https://doi.org/10.1080/17439760.2014.967802

Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education*, *40*(1), 457-496. https://doi.org/10.3102%2F0091732X16669319

Schwab, A. (2012). Epistemic humility and medical practice: Translating epistemic categories into ethical obligations. *Journal of Medicine and Philosophy*, *37*(1), 28-48. https://doi.org/10.1093/jmp/jhr054

Schei, E., Fuks, A., & Boudreau, J. D. (2019). Reflection in medical education: intellectual humility, discovery, and know-how. *Medicine, Health Care and Philosophy*, *22*(2), 167-178. https://doi.org/10.1007/s11019-018-9878-2

Sinatra, G. M. (2016). Thoughts on knowledge about thinking about knowledge. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of Epistemic Cognition*, 479–491. https://doi.org/10.4324/9781315795225

Stanovich, K. E., & Toplak, M. E. (2019). The need for intellectual diversity in psychological science: Our own studies of actively open-minded thinking as a case study. *Cognition*, *187*, 156-166. https://doi.org/10.1016/j.cognition.2019.03.006

Tanesini, A. (2018). Intellectual humility as attitude. *Philosophy and phenomenological research*, *96*(2), 399-420. https://doi.org/10.1111/phpr.12326

Tangney, J. P. (2000). Humility: Theoretical perspectives, empirical findings and directions for future research. *Journal of Social and Clinical Psychology*, *19*(1), 70-82. https://doi.org/10.1521/jscp.2000.19.1.70

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology*, *50*(4), 1037. https://doi.org/10.1037/a0034910

Van Tongeren, D. R., Davis, D. E., Hook, J. N., & Witvliet, C. V. O. (2019). Humility. *Current Directions in Psychological Science*, 28(5), 463-468. https://doi.org/10.1177%2F0963721419850153

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*(3), 506-519. https://doi.org/10.1037/a0028857

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one*, *11*(3), e0152719. https://doi.org/10.1371/journal.pone.0152719

Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, *94*(3), 509–539. https://doi.org/10.1111/phpr.12228

Zachry, C. E., Phan, L. V., Blackie, L. E., & Jayawickreme, E. (2018). Situation-based contingencies underlying wisdom-content manifestations: Examining intellectual humility in daily life. *The Journals of Gerontology: Series B*, *73*(8), 1404-1415. https://doi.org/10.1093/geronb/gby016

Zettler, I., Lang, J. W., Hülsheger, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process model to self-and observer reports. *Journal of Personality*, *84*(4), 461-472. https://doi.org/10.1111/jopy.12172

Zhang, H., Hook, J. N., Farrell, J. E., Mosher, D. K., Van Tongeren, D. R., & Davis, D. E. (2018). The effect of religious diversity on religious belonging and meaning: The role of intellectual humility. *Psychology of Religion and Spirituality*, *10*(1), 72. https://psycnet.apa.org/doi/10.1037/rel0000108

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

Table 1.1

Intellectual Humility Content Domains

Domain	Descriptions
Open to other ideas	Open to receiving new information or ideas, even when contrary. Open to revising one's viewpoint. Lack of rigidity in views. Greater likelihood of considering proposed alternative ideas.
Low ego interference	Regulation of one's concern for being right. Unthreatened by intellectual disagreement. Moderate reactions to differences of opinion. Absence of discomfort and negative emotion when considering limitations.
Awareness of limitations of beliefs	Insight regarding the limitations of one's knowledge and understanding. Recognition of the potential fallibility of beliefs. Acknowledgement of gaps in one's knowledge. Attending to the limitations in the evidence supporting beliefs.
Awareness of limitations in ability	An accurate view of intellectual strengths and weaknesses. Recognition of limitations in one's ability to obtain and evaluate pertinent information. Accepting that one's cognitive faculties are limited and imperfect.
Pursues truth / seeks evidence	Pursuing and incorporating knowledge and truth from various sources, including potentially disagreeable sources. Attending to the evidentiary basis of one's views. Being appropriately concerned with the way one acquires and applies knowledge. Desiring to gain knowledge and increase understanding.
Honest and fair	Presenting evidence for one's ideas fairly. Avoiding manipulative strategies for influencing others. Less likely to pretend to know something.
Relational effectiveness	Respectful and attentive to the perspective of others during discussions of differences. Sensitive to how ideas will be received by specific others. Willingness to compromise or negotiate. Relatively less interpersonal conflict due to ideas.
Low arrogance / overconfidence	Rarely assuming competence. Regulating arrogance. Avoiding unjustified confidence in the evidentiary basis for one's beliefs. Lack of overconfidence regarding the correctness of opinions and beliefs. Less likely to base a confident decision on incorrect information. Tends not to pretend to have knowledge.
Valuing the knowledge of others	Work collaboratively with those with more informative perspectives. Respect for and openness to the views of other people.

Table 1.2

Basic Properties of Intellectual Humility Measures

Scale name	Example item	Number of factors	Alpha	Temporal stability	Convergent evidence (|r| > .30)	Criterion evidence
Cultural Humility Scale (Hook et al., 2013; 12 items); Other-report	Is genuinely interested in learning more.	2	Full scale from .86 to .93; Subscale from .84 to .93	Not tested	Not tested	retrospective report of working alliance; report of ongoing working alliance; report of ongoing working alliance;
Intellectual Humility Scale (McElroy et al., 2014; 16 items); Other-report	Always has to have the last word in an argument. (reverse coded)	2	Full scale of .94 and .96; Subscale from .92 to .94	Not tested	Positively correlated with trust, agreeableness, openness, conscientiousness, and benevolent transgression-related interpersonal motivation. Negatively correlated with neuroticism, anger towards God, and avoidant transgression-related interpersonal motivation.	IHS scores predict rater-selected exemplars of extreme high or low IH
Specific Intellectual Humility Scale (Hoyle et al., 2016; 9 and 3 item versions); Self-report	When it comes to my views about _____ I may be overlooking evidence.	1	From .88 to .96	Not tested	For some content topics, positively correlated with general intellectual humility.	evaluation of forms of evidence with respect to a specific topic (for few topics and forms of evidence only despite many tested)
Comprehensive Intellectual Humility Scale (Krumrei-Mancuso et al., 2016; 22 items); Self-report	I can respect others, even if I disagree with them in important ways.	4	Full scale from .82 to 88; Subscale from .70 to .89	Full scale of 75 (1 month) and .70 (3 months); Subscale from .56 to .74 (1 month) and .50 to .76 (3 months)	Positively correlated with intellectual openness, self-correction, accurate self-perception, open-minded thinking, and openness to experience. Negatively correlated with intellectual arrogance.	Open-minded thinking and openness to experience when controlling for other relevant variables
General Intellectual Humility Scale (Leary et al., 2017; 6 items); Self-report	I accept that my beliefs and attitudes may be wrong.	1	From .73 to .87.	Not tested	Positively correlated with openness epistemic curiosity, existential quest, need for cognition, intellectual humility, open-mindedness, autonomy, and courage. Negatively correlated with dogmatism, intolerance of ambiguity, and self-righteousness.	More moderate religious views; preference for balanced arguments; less negative response to disagreeable information; more favorable attitudes toward those who change their mind; greater discernment of argument quality
Limitations-Owning Intellectual Humility Scale (Haggard et al., 2018; 12 items); Self-report	I care about the truth.	3	full scale: .87; subscales: .80-.87	.75 (5 months)	Positively correlated with intellectual humility, agreeableness, conscientiousness, neuroticism. Negatively correlated with neuroticism and closed-mindedness.	closed-mindedness, assertiveness, authentic pride

Scale name	Example item	Number of factors	Alpha	Temporal stability	Convergent evidence (|r| > .30)	Criterion evidence
Porter and Schumann Intellectual Humility Scale (PSIHS; Porter & Schumann, 2018; 9 items); Self-report	I am willing to admit it if I don't know something.	1	Average of .74	Not tested	Positively correlated with general humility, agreeableness, growth mindset, openness to experience, need for cognition, modesty, learning goals, and conscientiousness.	Attributions of respectful intent and greater openness in response to disagreement scenarios. Preference for reading about an opposing view rather than a congenial view.
Alfano Multi-Dimensional Intellectual Humility Scale (AMIHS; Alfano et al., 2017; 22 items); Self-report	If someone points out an intellectual mistake that I've made, I tend to get angry. (reverse coded)	4	Not reported (item-response theoretic analyses presented instead)	Not tested	Positively correlated with informant reports. Negatively correlated with narcissism at the subscale level.	Open-mindedness subscale predicting overclaiming bias most strongly in regression with Big Six personality traits.
Theistic Intellectual Humility Scale (TIHS; Hill et al., 2021; 11 items); Self-report	I try to submit all my intellectual efforts to God	3	Full scale .83-.85	Not tested	Positively correlated at the subscale level with Conscientiousness and Need for Cognition. Negatively associated at the subscale level with hubristic pride.	Favorably moderating the influence of religious commitment on several indicators of well-being.
State-Trait Intellectual Humility Scale (STIHS; Zachry et al., 2018)	I search actively for reasons why my beliefs might be wrong	1	Full scale .91	Not tested	Positively correlated with authentic pride, Extraversion, Agreeableness, Conscientiousness, Openness, Altruism, prestige, existential quest, actively open-minded thinking, and tolerance. Negatively correlated with dogmatism, self-righteousness, and need for closure.	Aggregated state measure scores taken from experience sampling.

2 THE VALIDITY OF THE GENERAL INTELLECTUAL HUMILITY SCALE AS A MEAURE OF INTELLECTUAL HUMILITY

When people fail to recognize the limitations of their own perspectives, systems that rely on the exchange of ideas can break down, contributing to a variety of societal problems. Bias and arrogance can undermine crucial decision-making processes (Hiller & Hambrick, 2005), political polarization can shut down a nation's politics (Garimella & Weber, 2017; Lee et al., 2018; Levendusky & Malhotra, 2016), and holding a view based on misinformation can lead to dangerous public health situations (Enders, et al., 2020; Joslyn, 2017; Marlon et al., 2016). These and other situations might be benefited by a deeper understanding of intellectual humility, and in the past decade, psychologists and other scholars have collaborated to begin the empirical investigation of this virtue.

Early work in the empirical study of trait intellectual humility resulted in the development of many self-report questionnaire measures, and these questionnaires have been the most common method employed in empirical studies of intellectual humility (McElroy-Heltzel et al., 2019; Porter et al., 2021). Though direct self-report is a widely employed method of assessment in the social sciences, it is not appropriate for every application (Paulhus & Vazire, 2007). Some scholars have noted that intellectual humility and closely-related constructs seems to be especially bad candidates for the use of self-report measures (Christen et al., 2019; Brienza et al., 2018), and early in the course of studying intellectual humility, multiple studies have found evidence suggesting that self-report may not be accurate (Alfano et al., 2017; Haggard et al., 2018; Hill et al., 2021; Meagher et al., 2015; Meagher et al., 2021; Meagher, 2022; Zachry et al., 2019).

Though the problem of desirability bias has received the most attention in discussions of measuring intellectual humility, a potentially more serious threat is that it seems intellectual humility is required in order to provide an accurate self-report of intellectual humility. Specifically, similarly to how people sometimes lack the perspective required to recognize their limitations with respect to cognitive tasks (Kruger & Dunning, 1999; Pennycook et al., 2017), people low in intellectual humility may simply be unaware of their limitations with respect to intellectual humility. In contrast, those highest in intellectual humility may be attentive enough to universal human limitations that they might rate themselves moderately without any intent of downplaying their strengths. This would mean that self-reports may be perversely related with intellectual humility among people who strongly value qualities associated with intellectual humility, even in the absence of any form of desirability bias.

Because initial validity testing tends to be generally confirmatory (Maul, 2017), problems with measurement can easily slip past initial validity testing (Borsboom, 2006), and more critical validity testing conducted from outside of the author team is generally required to identify and probe potential problems (Kane, 2006). In the case of self-report measures of intellectual humility, the threat of inaccurate self-report is a serious one in need of additional investigation. The purpose of the present study is to conduct a set of more critical tests of validity for one of the more promising self-report measures of intellectual humility and, in doing so, gather evidence about the accuracy of the self-report of intellectual humility as well.

Intellectual Humility as a Psychological Construct

The intellectual humility concept in psychology has its origins in the philosophy literature on virtue epistemology, an area of philosophy that considers the possession of strong intellectual character as the means by which individuals are able to arrive at true beliefs and other epistemic

goods (Church & Samuelson, 2017). As one of the epistemic virtues, philosophers have defined intellectual humility variously, but a definition provided by Whitcomb et al. (2017) has been particularly influential on early psychological measurement work, having been cited as either influential or foundational for several instruments. Whitcomb et al. define intellectual humility as "proper attentiveness to, and owning of, one's intellectual limitations" (2017, p. 520) and argue that intellectual humility in practice emerges from dispositions spanning cognition, emotion, and behavior. Notably, the limitations-owning account does not necessarily require a person to be free of arrogance in order to be intellectually humble, but it is argued that a rational person who is intellectually humble will also be lacking in arrogance (Whitcomb et al., 2017, pp. 20-25).

In translating the intellectual humility (IH) concept from philosophy into psychological research, psychologists have provided numerous definitions alongside the introduction of new measurement instruments. The earliest published measurement instruments defined IH as a subdomain of general humility (Hook et al., 2015; McElroy et al., 2014). On this account, IH "involves an accurate or moderate view of one's strengths and weaknesses" concerning knowledge and intellectual influence along with "being interpersonally other-oriented rather than self-focused, marked by an ability to restrain egotism" (McElroy et al., p. 20). Subsequent authors of IH measures have attempted to specify the nature of intellectual humility independently rather than characterizing it chiefly in terms of general humility, though some of these authors nevertheless grant that intellectual humility can be considered a subdomain of general humility (Porter & Schumann, 2018). Measure authors have defined intellectual humility as "a nonthreatening awareness of intellectual fallibility" (Krumrei-Mancuso & Rouse, 2016, p. 210), "the recognition that a particular personal view may be fallible accompanied by appropriate attentiveness to limitations in the evidentiary basis of that view and to one's own limitations in

obtaining and evaluating information relevant to it" (Leary et al., 2017, p. 1), "the mean between intellectual arrogance and intellectual servility" (Haggard et al., 2018, p. 185), and as "a willingness to recognize the limits of one's knowledge and appreciate others' intellectual strengths" (Porter & Schumann, 2018, p. 140).

Though these definitions show a breadth and diversity in what authors have centralized in their conceptualization of IH, these authors' expositions of the IH concept clearly reflect a shared influence of virtue epistemology on the understanding of the construct. Those high in IH are expected to "receive contrary ideas without taking offense" and to handle conflicting views without being impatient or defensive (McElroy et al., 2014, p. 20). When they lack justified knowledge, they should "be receptive to inviting multiple perspectives and altering views if new information arises" (Haggard et al., 2018, p. 185). They are expected to "pay greater attention to evidence that bears on their beliefs" (Leary et al., 2017, p. 13) and to "be open to alternate evidence" even when it is contrary to their existing views (Krumrei-Mancuso & Rouse, 2016, p. 210).

In summary, psychological conceptualizations of IH have tended to incorporate the idea of a generally humble person with notions of intellectual virtue. The claim of intellectual or epistemic virtue is present in the conceptualization of IH across measures generally, and it is this claim that invites the most serious complications for self-report. Accordingly, a priority for the intellectual humility measurement literature is testing the epistemic expectations of a measure focused on this aspect of intellectual humility.

Development and Initial Validity Evidence for the GIHS

The present study focuses upon on the General Intellectual Humility Scale (GIHS; Leary et al., 2017). The GIHS was selected for the present study because it is the only measure of

intellectual humility entirely focused on cognition and metacognition, which are the aspects of intellectual humility that are most in need of additional testing with respect to potential problems with self-report. The authors of the GIHS define intellectual humility as "recognizing that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in obtaining and evaluating relevant information" (Leary et al., 2017, p. 793). Additionally, because these epistemic elements of intellectual humility are what most clearly distinguishes the construct from general humility, testing this element in isolation is likely to provide clearly complementary insights to findings from prior work examining the accuracy of self-reported humble character (e.g., Zettler et al., 2016).

In terms of situating the measure within a nomological network, GIHS scores were shown to be positively correlated with several other relevant constructs, including openness (r = .33), epistemic curiosity (r = .35), existential quest (r = .35), need for cognition (r = .34), a single-item self-report of intellectual humility (r =.37), open-mindedness (r =.43), intellectual autonomy (r = .31), and intellectual courage (r = .30; Leary et al., 2017). Scores were also shown to be negatively correlated with dogmatism (r = -.49), intolerance of ambiguity (r = -.32), and self-righteousness (r = -.35). These findings are generally in line with what one would expect from theory, showing sizable correlations with other intellectual virtues, openness, and cognitive dispositions compatible with intellectual humility. However, the most informative single result of these correlations is the especially large negative relationship with dogmatism, which raises the question of the extent to which the GIHS reflects something distinct from low dogmatism. An additional limitation of this evidence is the omission of comparisons with cognitive variables

such as cognitive reflection (Frederick, 2005), an instrument that not only overlaps conceptually but which has the benefit of objective scoring.

In addition to situating their measure within a nomological network of related constructs, the authors also offered initial criterion-related evidence of validity in the form of three studies. In the first of these three studies, the authors examined a set of predictions concerning how one reacts when faced with a position with which one disagrees. The primary claims to be tested were (a) that people scoring high on the GIHS "should be more willing to entertain beliefs that differ from their own" and (b) that they would "judge people whose views differ from theirs less negatively" (Leary et al., p. 798). However, no analysis was presented showing the relationship between GIHS scores, specific beliefs held by participants, and the reaction of participants to beliefs that differ from their own or people holding differing beliefs. Instead, the authors showed that the GIHS predicted lesser belief certainty, lesser belief extremity, greater agreement with arguments that religion has negative impacts on society, and a general tendency to favorably rate essay content and author characteristics regardless of the essay's tone toward religion. The finding that GIHS scores predicted general agreement with the content of every essay condition is especially noteworthy in that it precludes treating any of the essays as disagreeable content for those scoring high on the GIHS. Thus, we do not have an answer concerning how those high on the GIHS react to beliefs that differ from their own nor how they judge people whose views differ from their own.

In the second study, the authors tested a prediction concerning reactions to people changing their attitude about a topic. Participants read a scenario about a fictional political candidate who had announced a change in his view on an unspecified environmental issue. The primary hypothesis was that people scoring high on the GIHS would rate the candidate who

changed his mind more positively than those with low scores on the GIHS. This hypothesis was partially supported, with interesting implications for the overall measurement claims of the GIHS. The GIHS did predict a markedly more favorable view toward the candidate changing his viewpoint among Republicans. However, scores on the GIHS did not result in more positive evaluations of the candidate among either Democrats or Independents, departing from the authors' prediction that the GIHS would generally predict more positive responses to a view change.

This final study sought to demonstrate that GIHS scores would predict greater sensitivity and responsiveness to argument quality about the importance of flossing, especially among individuals who did not report flossing regularly. Whereas the GIHS did predict greater sensitivity to argument quality among non-flossers, the GIHS did not predict greater sensitivity to evidence quality among those who flossed regularly. Because people in general may pay closer attention to arguments that are challenging to existing beliefs or behaviors and which may result in needing to act (Ditto et al., 1998), it is not clear that this result should be interpreted as indicative of IH. Instead, if the GIHS had predicted greater sensitivity to argument quality even for information that was congenial to existing beliefs and personally flattering, that would have been more congruent with the expectation that those high in IH are distinctively attentive to the evidentiary basis of their own beliefs.

Altogether, these studies (Leary et al., 2017) demonstrate that GIHS scores predict a degree of intellectual agreeableness and engagement that may be consistent with IH, but they do not establish the core claim in the GIHS IH definition: consideration that one's beliefs might be fallible involving attentiveness to the evidentiary basis of one's own beliefs and to one's own limitations in assessing such evidence. If GIHS scores are reflective of qualities related to IH

such as valuing knowledge and critical thought rather than being reflective of IH per se, this is a reason for concern because those qualities may at times contribute to arrogance (e.g., Gibson, 2003). Given these ambiguities, it important to continue to study the GIHS through a second wave of more pointedly critical tests of construct validity.

The Present Study

The principal shortcoming of the original validity paper for the GIHS (Leary et al., 2017) was that it did not directly test the core claims of the instrument–that those scoring high on the GIHS are strong at attending to the limitations in the evidentiary basis of their beliefs and to their own limitations with respect to assessing the quality of evidence pertaining to their beliefs. In the interest of more decisively determining the strengths and limitations of the GIHS as a measure of IH, I identified and conducted three theory-informed tests of these core claims.

Test 1: GIHS Scores and Epistemically Unwarranted Beliefs

Leary et al. (2017) contended that people high on the GIHS are aware of the evidentiary basis of their beliefs and noted that this could be expected to lead to holding beliefs that are more realistic over time. One way to more directly test this prediction is to see how well GIHS scores predict endorsement of beliefs that are generally accepted as unwarranted or false (e.g., the moon landing was a hoax, vaccines cause autism, ghosts haunt old buildings). If a person endorses many such unwarranted beliefs, that would suggest that they are not especially attentive to the evidentiary basis of their beliefs, which should be reflected in a lower score on the GIHS. Even more so, if a person expresses certainty that such beliefs are true or persists in believing in such beliefs despite claiming to have researched the issue, then that would demonstrate a lack of attentiveness to evidence and lack of awareness of one's limitations in evaluating evidence, both qualities that would be strongly in conflict with the GIHS notion of IH. The GIHS should thus be

strongly predictive of lesser endorsement of false beliefs, lesser certainty about false beliefs, and lesser tendency to have claimed to research a topic while endorsing the unwarranted view as true.

Test 2: GIHS Scores and Bias Blind Spot

In addition to attending to the evidence supporting specific beliefs, the GIHS is also meant to reflect awareness of "one's own limitations in obtaining and evaluating relevant information" (Leary et al., 2017, p. 793). Initial validity tests did not directly test this claim, yet one way of testing this would be to examine how well GIHS scores correlate with a measure of bias blind spot (Pronin et al., 2002). The bias blind spot refers to the tendency for people to think that bias is a problem for other people but is problematic to a lesser degree for themselves. Cognitive bias blind spot tests specifically check if someone is self-aware that they are limited by certain robust intellectual biases, and people who are otherwise cognitively sophisticated remain susceptible to a cognitive bias blind spot and may even be more prone to that blind spot (West et al., 2012). If those scoring high on the GIHS are aware of their own limitations with respect to engaging and evaluating information, then that would necessarily include an awareness and acceptance that they are susceptible to universal cognitive biases. If, on the other hand, the GIHS is associated with increased bias blind spot, then that would raise the question of whether the endorsements of GIHS items are similarly rooted in excessively positive self-evaluation. Accordingly, GIHS scores should be negatively associated with bias blind spot.

Test 3: GIHS Scores and Political Polarization about Global Warming

The GIHS authors contended that people high in IH should not only be generally attentive to evidence but should specifically be receptive to evidence that conflicts with existing beliefs (Leary et al., 2017). One way to provide a rigorous test of this prediction is to test whether an individual aligns with an evidence-based view on topics even when cultural and

group affiliations provide obstacles to accepting that view. A strong example of the difficulty of navigating conflicts between evidence, identity, and group membership can be seen in motivated skepticism concerning anthropogenic climate change. Political ideology and political affiliations strongly influence beliefs about global warming despite a clear scientific consensus on the matter and despite efforts to educate the public on relevant matters. Importantly, cognitive strengths such as greater education levels, intelligence, effortful thinking, and critical thinking amplify political polarization rather than protect against it, with these strengths allowing for more sophisticated motivated reasoning (Kahan, 2015). This is likely because skepticism concerning anthropogenic global warming is largely driven by personal motivations rather than consideration of evidence (Nisbet & Scheufele, 2009). This makes acceptance of anthropogenic climate change among American conservatives an apt test for the GIHS because it represents a case where the evidence supporting a position is clear yet is often rejected either without critical reflection or through motivated reasoning. Accordingly, if the GIHS scores attenuate the effect of political orientation on belief in anthropogenic global warming and predict belief in anthropogenic global warming among conservatives, it would be a strong point of evidence in support of the construct validity of the GIHS.

In summary, the purpose of the present study was to conduct three sets of tests of validity in order to advance the understanding of the GIHS and self-report of IH generally. To the degree that GIHS scores reflect IH as described by Leary et al.'s (2017) definition of IH, the following hypotheses should hold.

RQ1: GIHS Scores and Epistemically Unwarranted Beliefs

H1a: GIHS scores should correlate negatively with endorsement of unwarranted beliefs.

H1b: GIHS scores should correlate negatively with expressing certainty that an unwarranted belief is true.

H1c: GIHS scores should correlate negatively with claiming to have carefully researched an unwarranted belief yet still endorsing it as true.

RQ2: GIHS Scores and Bias Blind Spot

H2: GIHS scores should correlate negatively with cognitive bias blind spot.

RQ3: Do those scoring high on the GIHS accept evidence-supported views when doing so is difficult?

H3a: GIHS scores should attenuate the influence of political orientation on belief in anthropogenic global warming.

H3b: GIHS scores should be positively associated with belief in anthropogenic global warming among political conservatives.

Method

Participants and Procedure

Participants (N = 481) were recruited from the research participant crowdsourcing platform Prolific (224 female [46.6%]; 246 male [51.1%]; 10 non-binary [2.1%]), ranging in age from 18 to 77 (M = 34.31, SD = 12.96). Prolific is a research-focused competitor of Amazon's widely used Mechanical Turk, which has recently come under scrutiny related to problems with data quality (Berinsky et al., 2012; Buhrmester et al., 2011; Casler et al., 2013; Goodman et al., 2013). Prolific has several advantages for data quality relative to Amazon's Mechanical Turk, Qualtrics, or other competitors (Peer et al., 2017). Perhaps most importantly, Prolific takes measures to maintain the quality of the respondent pool, promote accuracy in demographic information, and protect against some of the pitfalls that tend to plague competing platforms

(Palan & Schitter, 2018). As a major advantage compared to traditional survey firms such as Qualtrics, Prolific offers economical prescreening options, allowing for quota sampling to ensure greater diversity of beliefs and values relevant to my central research questions. Accordingly, in line with recent polling data from Gallup showing the United States adult population to be 35% conservative, 35% moderate, 26% liberal, and 4% other (Saad, 2019), I prescreened using Prolific Academic's demographic data to select 175 conservatives, 175 moderates, 130 liberals, and 20 identifying as "other" from the pool of American participants available through Prolific.

Recruitment and data collection occurred on March 10, 2021. Following informed consent, participants completed a questionnaire including demographic items and study measures. A total of 500 participants completed the study. Thirteen participants were excluded because they failed at least one of two manipulation checks, items which asked participants to input a specified response to show adequate attention to the task. An additional six participants were disqualified for not responding to any open response items, indicating inadequate effort. This resulted in the final sample of 481 participants.

The sample was predominately White/Caucasian, with 69.6% (n = 335) identifying as White/Caucasian, 9.6% (n = 46) as Asian/Pacific Islander, 6.2% (n = 30) as Black/African-American, 6.2% (n = 30) as Hispanic/Latino/a, 7.2% (n = 35) as multiracial/two or more races, and 1% (n = 5) as other. Regarding religious affiliation, 29.3% (n = 141) reported as Protestant Christian, 19.8% (n = 95) as Agnostic, 14.1% (n = 68) as Roman Catholic, 13.5% (n = 65) as Atheist, 11.2% (n = 54) as having no particular affiliation, 1.7% (n = 8) as Mormon, 1.7% (n = 8) as Buddhist, .8% (n = 4) as Jewish, .6% (n = 3) as Hindu, .4% (n = 2) as Orthodox Christian, 5.4% (n = 26) as other, and .4% (n = 2) provided no response. In addition to these affiliations, 22.7% (n = 109) indicated that they would also describe themselves as a born-again or

evangelical Christian, and 56.8% (*n* = 273) reported having a relationship with God or a higher being. With respect to the highest level of school completed or highest degree received, 1% (*n* = 5) reported less education than a high school degree, 12.7% (*n* = 61) reported a high school diploma, 24.9% (*n* = 120) reported some college, 8.1% (*n* = 39) reported an Associate's degree, 36.2% (*n* = 174) reported a Bachelor's degree, 14.1% (*n* = 68) reported a Master's degree, 1.5% (*n* = 7) reported a Doctorate, and 1.5% (*n* = 7) reported a professional degree.

Measures

General Intellectual Humility

General intellectual humility was measured using the General Intellectual Humility Scale (GIHS; Leary et al., 2017). The GIHS is a single-factor measure comprised of six items such as "I accept that my beliefs and attitudes may be wrong." Response choices are arranged on a 5-point Likert-like scale ranging from *not at all like me* to *very like me*, with higher scores intended to reflect greater general intellectual humility. Cronbach's coefficient alpha values for GIHS scores ranged from .73 to .87 across six samples. Additional validity evidence is discussed at length above.

Unwarranted Beliefs

The tendency to hold unwarranted beliefs and the tendency to claim certainty about such beliefs were measured by the Inventory of Epistemically Unwarranted Beliefs (IEUB; Dyer & Hall, 2018). The IEUB consists of 50 items with response choices arranged on a 5-point Likert-like scale, ranging from 1 (*totally sure it's false*) to 5 (*totally sure it's true*). Items consist of statements of claims that lack adequate supporting evidence (e.g., routine childhood vaccines cause developmental problems such as autism in children). The scale is comprised of such statements from six content categories to provide broad coverage of a variety of beliefs:

paranormal, religious, health pseudoscience, extraordinary life forms, conspiracy theories, and ghosts. An unscored subscale of 12 general science knowledge items is also included to prevent strategic responding (i.e., recognizing that all items will be false). Higher subscale scores reflect holding unwarranted beliefs in a specific content domain and higher total scores reflect holding unwarranted beliefs generally. At the item-level, scores of four or five (after reversing where appropriate) reflect endorsing a belief and scores of five reflect endorsing a belief as certainly true, so belief and belief certainty can be assessed by dichotomizing scores at these points (Dyer & Hall, 2018, p. 300).

Confirmatory principal components analysis has supported the 6-category subscale structure of the IEUB. IEUB scores were shown to be stable over time, with average scores in a control group showing no statistically significant change in score between testing at the start and end of the semester for total score and for all 6 subscale scores. The IEUB also showed strong criterion-related evidence of validity in reflecting the effect of instruction in critical thinking and pseudoscience on scale scores over a semester. The IEUB showed additional evidence of validity in demonstrating expected relationships with other variables pre-intervention including academic major and self-reported SAT scores.

The tendency to endorse unwarranted beliefs despite claiming to have researched the issue was also assessed using the IEUB. After completing the IEUB, participants were presented with a list of the unwarranted beliefs they endorsed (scores of four or five), generated via logic operations native to the Qualtrics survey platform. Participants were asked to select any beliefs that they had carefully researched to determine the truth of the issue. The number of selections on this screen determined the count of endorsed beliefs also endorsed as carefully researched. In conjunction with IEUB results, this additional item set allows for identifying cases where a

respondent has endorsed an unwarranted belief despite claiming to have carefully researched the topic, suggesting selective attention to evidence, overestimation of one's ability to evaluate the relevant information, over-claiming about what has been researched, or some combination of these.

Belief in Anthropogenic Global Warming

Belief in anthropogenic global warming was assessed by an unnamed measure created by Hennes et al. (BAGW; 2016). The BAGW is a single-factor scale comprised of seven items such as "Do you believe that global warming is anthropogenic (caused by human behavior)?" Response choices are arranged on a 7-point Likert-like scale ranging from 1 (*definitely not; not at all likely*) to 7 (*definitely, extremely likely*). High scores reflect strong belief in anthropogenic climate change while low scores reflect strong denial of anthropogenic climate change. This scale showed strong evidence of reliability with Cronbach's coefficient alpha of .96 and convergent evidence of validity in demonstrating expected relationships with attitudes toward the environment, economic values, and political alignment (Hennes et al., 2016).

Science Intelligence

Scientific intelligence was assessed by The Ordinary Science Intelligence Scale 2.0 (OSI; Kahan, 2017). The OSI is single-factor measure comprised of 18 items. It is designed to "measure a latent capacity to recognize and make use of valid scientific evidence in everyday decision-making" (Kahan, 2017). Though it is a novel combination of elements, it is composed from prior measures: science knowledge from the NSF Indicators Battery (National Science Board, 2014) and the Pew Science and Technology battery (Pew Research Center, 2013), understanding of scientific methods from the NSF Indicators Battery, quantitative reasoning ability from the Lipkus/Peters Numeracy battery (Lipkus et al., 2001), and a 3-item test of

cognitive reflection (Frederick, 2005). The full scale has shown strong evidence of reliability, with initial testing showing a Cronbach's alpha of .83. Factor analysis of the OSI supported the claim of a general factor accounting for observed variance among all items, and analysis of item response curves demonstrated strong evidence of reliability across the range of item difficulty (Kahan, 2017). Additionally, the OSI also showed strong criterion-related evidence of validity in demonstrating expected relationships with other cognitive constructs including numeracy, education level, and actively open-minded thinking (Kahan, 2017).

Political Orientation

Following prior research demonstrating the influence of political affiliation and ideology on the exercise of critical thinking about scientific facts (Kahan, 2016), political party self-identification was measured with a 7-point scale (strong Democrat, Democrat, independent lean Democrat, Independent, independent lean Republican, Republican, strong Republican), and political ideology was measured with a 5-point item (very liberal, liberal, moderate, conservative, very conservative). A total left-right political orientation score was derived by summing the distance from the median point for both scales.

Bias Blind Spot

Bias blind spot was measured using the Bias Blind Spot Questionnaire (BSQ; West et al., 2012). The BSQ is a single-factor assessment comprised of 7 items. Response choices are arrayed on a 5-point Likert-like scale with response choices ranging from 1 (*not at all likely*) to 5 (*very likely*). The BSQ reflects the extent to which people believe themselves to be superior to others at avoiding biases which have been demonstrated to be universal. Higher scores reflect excessive confidence concerning one's ability to avoid cognitive bias. Respondents are presented with a description of each cognitive bias and asked to rate how susceptible they believe

themselves to be and how susceptible they believe the average person in their context (e.g., average student at a specific school, average American) to be. Rating others as more likely than oneself to have these biases is therefore considered an inaccurate assessment of one's own limitations relative to the limitations of others because the included biases have been shown to be intractable to superior intellectual ability or cognitive sophistication. Examples of specific biases represented in the questionnaire include myside bias, framing effects, and anchoring effects. Overall scores for bias blind spot are obtained by summing the differences between self-rating and the rating of other people for each item. The BSQ has demonstrated criterion-related evidence of validity in that it was shown to be predictive of unwarranted confidence rather than a genuine lesser susceptibility to bias, confirmed by testing the ability of individuals and finding that those claiming to do better than average did not in fact do better (West et al., 2012).

Power Analysis

Sample size was determined based on an estimate of what would be needed to detect the predicted interaction effects. Based on prior research, the effect size of the relationship between political orientation and BAGW was expected to be large (e.g. Evans & Feng, 2013). Assuming a large effect size (f^2 = .35), an alpha of .05, a power of .8, and one predictor, G*Power 3.1 indicated a sample size of 25 would be needed to detect the effect of political orientation on BAGW (Faul et al, 2009). Because a two-way interaction involving GIHS and political orientation was expected to attenuate rather than eliminate or reverse the main effect, the sample size to detect an interaction reducing the effect by 50% would need to be fourteen times as large as was needed to detect the main effect with an alpha or .05 and a power of .80, indicating a required sample size of 350 (Giner-Sorrola, 2018; Simonsohn, 2014).

Results

Preliminary Analyses

Analyses were conducted using IMB SPSS Statistics for Windows, Version 27**.** Of the 481 participants remaining after exclusions, thirty-eight cases had some missing data. The range of missing data for these participants was between 1.1% and 2.2%. Overall missingness was less than .1%, no scale or subscale variable had more than .2% of data missing, and no item had more than .8% of data missing. Missing responses for OSI items were coded as incorrect answers, affecting 12 of 8,654 total OSI data points, with no more than two data points for any item, and no more than one data point for any participant. The Little's MCAR test was significant, indicating that data were not likely to be MCAR. Data were thus assumed to be MAR after inspecting patterns of missingness for each variable (Schafer & Graham, 2002; Schlomer, et al., 2010). Regardless of the mechanism of missingness, the impact on findings is expected to be negligible given the very low level of missing data (Bennett, 2001; Peng et al., 2006; Schafer, 1999). Given these findings, missing data were handled with the available item analysis method (Parent, 2013).

Hypotheses were tested using bivariate correlations and linear regression models. A variety of quantitative descriptive statistics were examined for all study variables including means, standard deviations, minimum/maximum values, and quartiles along with visual plots (histograms) to better understand the distributional properties of the variables. Visual inspection of data revealed significant non-normality for AGW along with expected departures from normality for IEUB count variables. Accordingly, both Pearson (parametric) and Spearman (non-parametric) correlations were examined to assess bivariate relations among variables.

Linear regression has four core assumptions: (a) the relationships between independent variables and the dependent variable are linear, (b) the residual variance is the same across values of independent variables (homoscedasticity), (c) observations are independent, and (d) residuals are normally distributed (Pituch & Stevens, 2016). These assumptions were empirically evaluated using a variety of approaches including residual histograms (normality of residuals), q-q and p-p plots (normality of residuals), and residual-by-predicted value scatter plots (linearity and homoscedasticity). Multicollinearity was assessed by the variance inflation factor (VIF), and residuals were inspected for potential influential outliers using the plots described above and with Cook's and Mahalanobis distance statistics.

Empirical checks suggested that assumptions for BAGW were not adequately met due to non-normality of residuals. Model fit criteria and assumption checks (e.g., residual plots) supported using linear regression with the reversed and log transformed version of BAGW (Tabachnick, et al., 2019, 7th ed., pp. 75-76). After applying these transformations, all empirical evaluations of linear regression model assumptions were adequately met and there were no concerns regarding multicollinearity or influential outliers.

Bivariate correlations and descriptive statistics for primary study variables are reported in Table 2.1. GIHS scores were negatively associated with reporting having a relationship with God (r = -.12, p = .009) and negatively associated with identifying as Protestant (r = -.12, p = .012). GIHS scores were positively associated with identifying as an Atheist (r = .14, p = .002) and were unrelated with identifying as Agnostic (r = .07, p = .152). GIHS scores were unrelated to age (r = -.07, p = .128), education level (r = .06, p = .226), or gender (r = -.070, p = .128). GIHS scores were negatively associated with left-right political orientation (r = -.25, p < .001),

meaning higher GIHS scores were associated with being more liberal/Democrat and less conservative/Republican.

Research Question 1

The first set of three hypotheses concerned the question of whether GIHS scores would be negatively associated with epistemically unwarranted beliefs. Hypothesis 1a was that GIHS scores would correlate negatively with endorsement of unwarranted beliefs. GIHS scores were negatively correlated with IEUB mean score (r = -.13, p = .005) but were unrelated to the count of unwarranted beliefs endorsed (r_s = -.09, p = .054). Thus, Hypothesis 1a was partially supported. Hypothesis 1b was that GIHS scores would correlate negatively with the count of unwarranted beliefs endorsed with certainty. GIHS scores were unrelated to the count of unwarranted beliefs endorsed with certainty (r_s = -.08, p = .715). Thus, Hypothesis 1b was not supported. Hypothesis 1c was that GIHS scores would correlate negatively with the count of unwarranted beliefs both endorsed and claimed to have been carefully researched. GIHS scores were unrelated to the count of unwarranted beliefs both endorsed and claimed to have been carefully researched (r_s = .01, p = .828). Thus, Hypothesis 1c was not supported.

Research Question 2

Hypothesis 2 was that GIHS scores would be negatively related to bias blind spot as measured by the BSQ. GIHS scores were correlated positively with BSQ scores (r = .17, p < .001). This relationship was larger in effect size than the relationship between BSQ and OSI (r = .11, p = .017) and remained when controlling for OSI (B = .16, t[478] = 3.65, p < .001), suggesting that GIHS scores predict overconfidence related to avoiding bias not solely attributable to intelligence. Thus, Hypothesis 2 was not supported.

Research Question 3

The final set of hypotheses concerned the relationship between GIHS scores and political polarization about global warming. Hypothesis 3a was that GIHS scores would moderate the influence of political orientation on belief in global warming. Moderation analyses were conducted according to the process described by Cohen (1968). GIHS scores did not significantly moderate the influence of political orientation on belief in global warming (B = -.02, t[477] = -.11, p = .91; Table 2.3). Thus, Hypothesis 3a was not supported. Hypothesis 3b was that GIHS scores would predict belief in global warming when controlling for political orientation (seen here and elsewhere as a negative relationship due to reversed scores). GIHS scores significantly predicted belief in global warming when controlling for political orientation (B = -.22, t(478) = 5.69, p < .001; Table 2.4). Thus, Hypothesis 3b was supported.

Exploratory Analyses

Additional analyses were conducted to further explore primary results. Given that GIHS scores were not associated with the count of beliefs endorsed and were negatively correlated with IEUB mean scores, follow-up analysis was conducted to probe whether GIHS scores predicted stronger rejection of unwarranted beliefs. This analysis showed that GIHS scores were associated with rejecting unwarranted beliefs as certainly false (r_s = .15, p = .001).

To investigate the possibility that GIHS scores could show different relationships with unwarranted beliefs depending on the content of the specific beliefs, analyses run for the full scale IEUB score were repeated at the subscale level. Subscales from the IEUB showed largely the same pattern of associations with GIHS scores as seen with full scale scores, though there were some differences. Mean scores were significantly negatively correlated with GIHS scores for the Health (r = -.13, p = .003) and Religion subscales (r = -.21, p < .001), and for the

Religion subscale, GIHS scores were significantly negatively correlated with both the count of items endorsed (r_s = -.12, p = .009) and the count of items endorsed as certain (r_s = -.12, p = .012; Table 2.2). Given these observed relationships between GIHS scores and the Religion subscale and in consideration of the possibility that, for some participants, Religion items may reflect central rather than incidental religious beliefs (K. Dyer, personal communication, February 18, 2019), planned full-scale analyses were rerun with the Religion subscale removed to examine the sensitivity of findings to this potential source of error. With the Religion subscale removed, relationships between revised IEUB scores (IEUBr1) and GIHS scores were essentially similar. For IEUBr1, GIHS scores were also negatively associated with mean scores (r = -.09, p = .043) and positively associated with rejecting unwarranted beliefs as certainly false (r_s = .12, p = .007). For IEUBr1, GIHS scores were not associated with the count of beliefs endorsed as true (r_s = -.07, p = .117), were not associated with the count of beliefs endorsed as certainly true (r_s = .04, p = .349), and were not associated with the count of beliefs endorsed and claimed as carefully researched (r_s = .01, p = .828). Altogether, these findings suggest that the results of the primary analyses are reflective of the relationship between GIHS scores and susceptibility to epistemically unwarranted belief in general and not solely a function of any confounding bias introduced by the Religion subscale.

Because GIHS scores were shown to predict belief in global warming, a follow-up exploratory analysis was run to examine whether GIHS scores may moderate the influence of political orientation on belief in global warming for those high in science intelligence, a phenomenon that would be reflective of motivated reasoning (Kahan, et al., 2016). This analysis was initially indeterminate due to the lack of an observed motivated reasoning effect (two-way interaction between OSI and political orientation) either in a model including OSI and political

orientation (B = .15, t[474] = 1.00, p = .32) or in two-way and three-way models including GIHS (Table 2.5). However, because the narrower question of whether it is human activity that is responsible for global warming is more polarizing than whether warming is occurring (McCright et al., 2011), I suspected that the single item of the BAGW addressing the cause of global warming (BAGW1) would be more likley to reflect motivated reasoning. After confirming that BAGW1 was the item most strongly correlated with political orientation in the study sample (r = -.51, p < .001; r_s = -.53, p < .001), exploratory analyses were run with BAGW1 as the dependent variable. As with the full scale, this single item was also reversed and log transformed to address non-normal residual errors, at which point all assumptions for regression were met. GIHS scores did not moderate the influence of political orientation on BAGW1, (B = -.12, t[477] = 0.58, p = .571; Table 2.6). GIHS scores were again significantly associated with BAGW1 when controlling for political orientation (B = -.13, t[477] = -3.39, p < .001, Table 2.7). In contrast with BAGW, for BAGW1, a motivated reasoning effect was observed, with a significant OSI-by-political orientation interaction consistent with a motivated reasoning effect (B = .36, t[474] = 2.29, p = .023; Table 2.8). Probing the OSI-by-political orientation interaction at the 16th and 84th percentile values for political orientation showed that for left (score of -3 or lower) and central political orientations (-3 < score < 3), there was not a significant association between OSI and disbelief in global warming (p values > .05). However, for those with right political orientation (score of 3 or higher), increased OSI was associated with increased disbelief in Global Warming (.12 < B < .17, p values $\leq$ .05; Figure 2.1). Though it was expected that GIHS scores should attenuate this effect in a three-way interaction, this was not observed (B = .09, t(473) = .09, p = .929; Table 2.8). Importantly, however, it should be noted that the present study

was underpowered to reliably detect such a three-way interaction effect, so the absence of an effect found here should be interpreted with caution.

Discussion

The purpose of the present study was to conduct a set of more critical validity tests for the GIHS. Tests were selected to be closely linked to theory, to be meaningfully difficult, and to probe open questions from prior validity studies. Specifically, I set out to test whether the GIHS predicted lesser likelihood of holding unwarranted beliefs, greater self-awareness concerning one's own epistemic limitations, and greater likelihood of aligning beliefs with evidence despite the influence of political polarization. The findings of exploratory analyses were also reported.

In general, GIHS scores showed a pattern of results that does not support the proposed interpretation of scores described its authors. The GIHS is described as reflecting a cognitive disposition involving greater sensitivity to the potential fallibility of specific beliefs accompanied by attentiveness to the evidentiary basis of beliefs and one's own limitations with respect to evaluating that evidence (Leary et al., 2017). It was expected that such a disposition would be clearly reflected in lesser endorsement of unwarranted beliefs such as conspiracy theories, pseudoscientific health claims, and paranormal beliefs. Instead, while GIHS scores showed a small association with stronger rejection of unwarranted beliefs, GIHS scores did not predict a lesser likelihood of endorsing unwarranted beliefs and did not predict a lesser likelihood of endorsing unwarranted beliefs as certainly true. This finding is consistent with prior research showing negligible or no significant relationship between GIHS scores and conspiratorial belief (Bowes et al., 2021) and showing no significant relationship or only a small negative correlation with specific forms of unwarranted belief (Bowes & Tasimi, 2021). The present study extends prior research by showing that scores also did not predict lesser certainty that unwarranted

beliefs are true, which has been argued to be more central to intellectual humility than just holding justified beliefs (Hazlett, 2012; Whitcomb et al, 2017). Additionally, GIHS scores did not predict lesser likelihood of endorsing an unwarranted belief despite claiming to have carefully researched the issue to determine the truth of the matter, suggesting that GIHS scores do not predict greater awareness of one's own limitations with respect to evaluating evidence pertaining to one's beliefs. Notably, these small and null effects stand in contrast with associations observed with science intelligence as measured by the OSI in this same sample, which robustly predicted lesser endorsement of unwarranted beliefs, lesser endorsement of unwarranted beliefs as certainly true, and lesser likelihood to endorse an unwarranted belief despite claiming to have researched the topic.

Another expectation tested in this study was that GIHS scores would lessen bias blind spot as measured with the BSQ. Contrary to this expectation, GIHS scores instead predicted greater bias blind spot with a larger effect size than any other variable included in the study, and this effect remained when controlling for cognitive sophistication in the form of science intelligence. This finding suggests that GIHS scores may partially reflect excessive confidence with respect to how well one manages one's limitations, at least with respect to avoiding cognitive bias (West et al., 2012). This is a noteworthy finding because prior research has found the GIHS to be free from overclaiming bias (Deffler et al., 2016) and self-enhancement bias (Haggard et al., 2018; Hill et al., 2021; Zachry et al., 2019), a problematic form of bias that has consistently been associated with other instruments (Alfano et al., 2017, Haggard et al., 2018; Hill et al., 2021; Zachry et al., 2019). The present study, however, shows that other forms of bias more relevant to the self-report of the epistemic aspects of intellectual humility may not be detected by general measures of desirable responding or self-enhancement.

The most interesting finding for the GIHS came in what was expected to be the most difficult test of the instrument – predicting belief in anthropogenic global warming among political conservatives and attenuating political polarization about global warming. Because GIHS scores are purported to reflect an orientation to evidence and because prior research has shown GIHS scores to be predictive of lesser self-reported political polarization (Bowes et al., 2020) and some IH-relevant outcomes among political conservatives (Leary et al., 2017), it was expected that GIHS scores would predict lesser political polarization about a well-known issue with a clear evidence-based position supported by broad scientific consensus. The present study found GIHS scores to be strongly associated with belief in anthropogenic global warming regardless of political orientation, but GIHS scores did not attenuate the influence of political orientation on belief in global warming. This combination of findings was surprising because it was assumed that any influence on belief in global warming among conservatives would result from lesser susceptibility to political polarization and motivated reasoning. Instead, GIHS scores predicted a similar increase in belief in global warming regardless of political orientation. Exploratory follow-up analyses were conducted to investigate whether this positive influence of GIHS scores on belief in global warming could attenuate a motivated reasoning effect. GIHS scores were not shown to attenuate motivated reasoning, but this finding should be interpreted with caution because the present study was underpowered for detecting the expected three-way moderation effect. Given that results involving other unwarranted beliefs do not support the GIHS as reflective of aligning beliefs with evidence, it seems that GIHS scores are likely associated with greater belief in global warming for reasons other than the careful consideration of evidence and rejection of politically motivated reasoning. One possibility is that GIHS scores reflect an aspect of identity – identifying as someone who defers to evidence. This idea is

consistent with the uniformly positive influence of GIHS scores on belief in global warming regardless of political orientation and is consistent with research suggesting that belief in global warming reflects personal factors rather than consideration of evidence (Nisbet & Scheufele, 2009). By contrast, if the effect were driven by greater attention to evidence, we might expect to see a much larger change from Republican/Conservatives and little change at all for Independents and Democrat/Liberals, with those high in IH generally converging at a similar level of belief regardless of political orientation.

Findings in this study are also relevant to discussions of the relationship between intellectual humility and religious belief. Whereas GIHS scores did not generally predict lesser endorsement of epistemically unwarranted beliefs, lesser endorsement of such beliefs as certain, or lesser tendency to endorse an unwarranted belief while also claiming to have researched the topic, GIHS scores were negatively associated with all three of these outcomes for the Religion subscale. In conjunction with the negative association between GIHS and reporting a relationship with God, the present study adds to the body of literature showing self-reported intellectual humility scores to be negatively related with variables connected to religious belief in a way that suggests potential problems with measurement (Hill et al., 2021). In addition, the present study is the first demonstration of self-reported intellectual humility being associated with being more likely to actively disbelieve religious ideas. This is seen in scores being associated with identifying as Atheist but not with identifying as Agnostic and in scores being associated with expressing certainty that particular religious beliefs are false. These findings perhaps substantiate the finding that GIHS scores did not predict greater acceptance of a pro-religion article, in contrast to the alternative explanation that the lack of that effect was a function of the sample (Leary et al., 2017). It seems likely that asking directly about willingness to change beliefs is a

potential source of bias that causes scores to skew toward non-belief given that religious individuals hold some beliefs as matters of faith that are not subject to ordinary processes of evaluation and revision (Stanovich et al., 2019).

Finally, as the primary purpose of this study, these findings have implications for the accuracy of self-report in the assessment of intellectual humility. Whereas those scoring highly on the GIHS self-report as managing their beliefs with a grounding in evidence, the findings from this study suggest that they are as likely as others to believe baseless ideas, to do so with certainty, and to hold such beliefs despite having reportedly investigated the issue at hand. The positive correlation between bias blind spot and GIHS scores also suggests a potentially inflated idea of how well one manages one's intellectual limitations, which would constitute a direct threat to the response process upon which self-report measurement of intellectual humility is predicated. Finally, while the correlation with belief in global warming would have been powerful evidence of intellectual humility had it been accompanied by strong evidence of aligning beliefs with evidence generally, instead that finding is suggestive of GIHS scores reflecting an aspect of identity rather than a cognitive disposition, consistent with the alternative interpretation of GIHS scores as reflective of values.

In summary, GIHS scores were here shown to produce noteworthy results with respect to political identity yet failed to demonstrate expected relationships with the more general outcomes of unwarranted beliefs and bias blind spot. While these findings are only the results from one study, they are relatively straightforward departures from what would be expected from theory, and it is appropriate to consider whether the limitations of the self-report method may be one cause of these discrepancies. If so, similar problems may exist for the accuracy-related subscales of other intellectual humility measures, and until greater clarity emerges through additional

testing, it seems prudent to attend to the possibility of inaccurate self-report when interpreting the scores of self-report intellectual humility measures.

Limitations

This study had several limitations. The use of a convenience sample likely introduced sampling bias relevant to the results of this study. In particular, participants were more educated than the general population, less likely to believe in God than the general population, less diverse in race and ethnicity than the general population, and more left on left-right political orientation than the general population. It also seems likely that participants were more media literate and more likely to value scientific research given that they self-selected to be participants on a crowdsourcing platform advertised as for scientific research. Though these were not qualities that were assessed, media literacy and attitudes toward science may well have some relationship to GIHS scores or other study variables. Additionally, despite the steps taken by Prolific to support data quality, it is harder to know that respondents are who they say they are with respect to nationality and harder know to that they are giving adequate effort to the task than it would be for an in-person study. Another limitation was inadequate sample size to reliably detect a three-way interaction involving GIHS, OSI, and political orientation. Though such an effect was not central to the predictions tested in this study, identifying and probing such an interaction could reveal effects relevant to the study hypotheses. Finally, a similar study done in collaboration with one or more GIHS authors would likely be stronger in every respect, especially in providing the GIHS the best opportunities for success in the study design and the strongest defense in analysis and interpretation.

Future Research

Future research on the GIHS should explore the underlying cause of observed effects involving political orientation. This would likely best be accomplished through gathering qualitative data about the politically relevant outcomes to provide insight into the reasons for departure from group norms. Future studies should also explore whether GIHS scores predict against-group trends in belief or cognition for liberals or if such effects only occur for conservatives. Namely, it would be important to see if GIHS scores promote greater alignment with evidence for topics where liberals tend not be aligned with evidence, such as with the safety of nuclear power, natural gas, and fracking (Nisbet et al., 2015) or overestimation of risk with respect to some aspects of Covid-19 at the time of writing this (Lopez et al., 2021). Additionally, future research should explore the possibility that GIHS scores and scores from subscales covering similar content from other measures reflect values or an aspect of identity. If identifying as a reasonable person has mild protective effects on some unwarranted beliefs and contributes somewhat to acceptance of facts even on highly polarized issues, this could have implications for intellectual humility and potentially practical applications as well.

More generally, future research should continue to investigate validity for measures of intellectual humility. Similar objective or observational tests of validity for informant-report could also be informative. For example, informant ratings or coder ratings on the GIHS might show a different pattern of results for unwarranted beliefs, bias blind spot, or political polarization as what was seen here. It may also be fruitful to experiment with small changes to the measure such as making items explicit statements of values to remove the perverse influence of intellectual humility, using sliders rather than Likert scales, or changing the instruction set.

Conclusion

In the scientific study of intellectual humility, it is important to be aware of and attend to the limitations of measurement instruments. For this reason, validity theorists have emphasized the importance of following initial validation efforts with more critical tests of validity, and this advice surely pertains to self-report measures of intellectual humility. By carefully attending to theory, clearly linking theory to meaningfully difficult predictions, and being careful with the interpretations of scores, self-report measures of intellectual humility measures can play an important role in the overall study of intellectual humility.

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (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

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25, 464–469. https://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Brienza, J. P., Kung, F. Y., Santos, H. C., Bobocel, D. R., & Grossmann, I. (2018). Wisdom, bias, and balance: Toward a process-sensitive measurement of wisdom-related cognition. *Journal of Personality and Social Psychology*, *115*(6), 1093. https://doi.org/10.1037/pspp0000171

Bowes, S. M., & Tasimi, A. (2021, August 17). Clarifying the Relations between Intellectual Humility and Misinformation Susceptibility. https://doi.org/10.31234/osf.io/x37c5

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425. https://doi.org/10.1007/s11336-006-1447-6

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177%2F1745691610393980

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral

testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Christen, M., Alfano, M., & Robinson, B. (2019). A cross-cultural assessment of the semantic dimensions of intellectual humility. *Ai & Society*, *34*(4), 785-801. https://doi.org/10.1007/s00146-017-0791-7

Church, I., & Samuelson, P. (2017). *Intellectual humility: An introduction to the philosophy and science*. Bloomsbury Publishing.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological bulletin*, *70*(6p1), 426.

Davis, D. E., Worthington Jr, E. L., & Hook, J. N. (2010). Humility: Review of measurement strategies and conceptualization as personality judgment. *The Journal of Positive Psychology*, *5*(4), 243–252. https://doi.org/10.1080/17439761003791672

Davis, D. E., McElroy, S. E., Rice, K. G., Choe, E., Westbrook, C., Hook, J. N., Van Tongeren, D. R., DeBlaere, C., Hill, P., Placares, V., & Worthington Jr, E. L. (2016). Is modesty a subdomain of humility? *The Journal of Positive Psychology*, *11*(4), 439-446. https://doi.org/10.1080/17439760.2015.1117130

Davis, D. E., Hook, J. N., & Hill, P. C. (2019). Seven challenges of an interdisciplinary project to measure intellectual humility. *Journal of Psychology and Christianity*, *38*(3), 148-156.

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

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75*(1), 53. https://dx.doi.org/10.1037/0022-3514.75.1.53

Dyer, K. D., & Hall, R. E. (2018). Effect of Critical Thinking Education on Epistemically Unwarranted Beliefs in College Students. *Research in Higher Education, 60*(3), 293-314. https://doi.org/10.1007/s11162-018-9513-3

Enders, A. M., Uscinski, J. E., Klofstad, C., & Stoler, J. (2020). The different forms of COVID-19 misinformation and their consequences. *The Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-48

Evans, J. H., & Feng, J. (2013). Conservative Protestantism and skepticism of scientists studying climate change. *Climatic Change*, *121*(4), 595–608. https://doi.org/10.1007/s10584-013-0946-6

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41, 1149-1160.

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

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological review*, *102*(4), 652-670. https://doi.org/10.1037/0033-295X.102.4.652

Garimella, V. R. K., & Weber, I. (2017). A long-term analysis of polarization on Twitter. Proceedings of The 11th International AAAI Conference on Web and Social Media (ICWSM): 528-531.

Gibson, J. M. (2003). Arrogance-a dangerous weapon of the physics trade? *Physics Today*, *56*(2), 54–55. https://doi.org/10.1063/1.1564350

Giner-Sorolla, R. (2018, January 24). Powering Your Interaction. Approaching Significance. https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Haggard, M., Rowatt, W. C., Leman, J. C., Meagher, B., Moore, C., Fergus, T., Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2018). Finding middle ground between intellectual arrogance and intellectual servility: Development and assessment of the limitations-owning intellectual humility scale. *Personality and Individual Differences*, *124*, 184–193. https://doi.org/10.1016/j.paid.2017.12.014

Hennes, E. P., Ruisch, B. C., Feygina, I., Monteiro, C. A., & Jost, J. T. (2016). Motivated recall in the service of the economic system: The case of anthropogenic climate change. *Journal of Experimental Psychology: General*, *145*(6), 755-771. https://doi.org/10.1037/xge0000148

Hill, P. C., Lewis Hall, M. E., Wang, D., & Decker, L. A. (2021). Theistic intellectual humility and well-being: does ideological context matter? *The Journal of Positive Psychology*, *16*(2), 155-167. https://doi.org/10.1080/17439760.2019.1689424

Hiller, N. J., & Hambrick, D. C. (2005). Conceptualizing executive hubris: The role of (hyper-) core self-evaluations in strategic decision-making. *Strategic Management Journal*, *26*(4), 297–319. https://doi.org/10.1002/smj.455

Hodge, A. S., Mosher, D. K., Davis, C. W., Captari, L. E., Hook, J. N., Davis, D. E., & Van Tongeren, D. R. (2020). Political Humility and Forgiveness of a Political Hurt or Offense. *Journal of Psychology and Theology*, *48*(2), 142–153. https://doi.org/10.1177%2F0091647120911112

Hook, J. N., Davis, D. E., Owen, J., Worthington Jr, E. L., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology*, *60*(3), 353. http://dx.doi.org/10.1037/a0032595

Hook, J., Davis, D., Van Tongeren, D., Hill, P., Worthington Jr, E., Farrell, J., & Dieke, P. (2015). Intellectual humility and forgiveness of religious leaders. *The Journal of Positive Psychology*, *10*(6), 499–506. https://doi.org/10.1080/17439760.2015.1004554

Hopkin, C. R., Hoyle, R. H., & Toner, K. (2014). Intellectual humility and reactions to opinions about religious beliefs. *Journal of Psychology and Theology*, *42*(1), 50-61. https://doi.org/10.1177%2F009164711404200106

Joslyn, M. R. (2017, July 1). These 3 things help make sense of the dangerous Minnesota measles outbreak. *The Washington Post*. https://www.washingtonpost.com/news/monkey-cage/wp/2017/06/01/these3-things-help-make-sense-of-the-dangerous-minnesota-measles-outbreak/

Kahan, D. M., Braman, D., Slovic, P., Gastil, J., & Cohen, G. L. (2007). The second national risk and culture study: Making sense of and making progress in the american culture war of fact. *GWU Legal Studies Research Paper*, (370), 08-26. http://dx.doi.org/10.2139/ssrn.1017189

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, *36*, 1–43. https://doi.org/10.1111/pops.12244

Kahan, D. M., & Corbin, J. C. (2016). A note on the perverse effects of actively open-minded thinking on climate-change polarization. *Research & Politics*, *3*(4), 2053168016676705.

Kahan, D. M. (2017). 'Ordinary science intelligence': A science-comprehension measure for study of risk and science communication, with notes on evolution and climate change. *Journal of Risk Research*, *20*(8), 995–1016. https://doi.org/10.1080/13669877.2016.1148067

Kane M. (2006). Validity. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). American Council on Education and Praeger.

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), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Krumrei-Mancuso, E. J., Haggard, M. C., LaBouff, J. P., & Rowatt, W. C. (2019). Links between intellectual humility and acquiring knowledge. *The Journal of Positive Psychology*, *15*(2), 1–16. 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

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 and Social Psychology Bulletin*, *43*(6), 793-813. https://doi.org/10.1177/0146167217697695

Lee, C., Shin, J., & Hong, A. (2018). Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in

South Korea. *Telematics and Informatics*, *35*(1), 245–254. https://doi.org/10.1016/j.tele.2017.11.005

Levendusky, M., & Malhotra, N. (2016). Does Media Coverage of Partisan Polarization Affect Political Attitudes? *Political Communication*, *33*(2), 283–301. https://doi.org/10.1080/10584609.2015.1038455

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177%2F0272989X0102100105

Lopes, L., Hamel, L., Sparks, G., Kirzinger, A., Kearney, A., Stokes, M., & Brodie, M. (2021, October 28). KFF. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-october-2021/

Marlon, J. R., Linden, S. van der, Howe, P. D., Leiserowitz, A., Woo, S. H. L., & Broad, K. (2018). Detecting local environmental change: The role of experience in shaping risk judgments about global warming. *Journal of Risk Research*, *22*(7), 1–15. https://doi.org/10.1080/13669877.2018.1430051

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 51-69. https://doi.org/10.1080/15366367.2017.1348108

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington Jr, E. L., & 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%2F009164711404200103

McElroy-Heltzel, S. E., Davis, D. E., DeBlaere, C., Worthington Jr, E. L., & Hook, J. N. (2019). Embarrassment of riches in the measurement of humility: A critical review of 22 measures. *The Journal of Positive Psychology*, 14(3), 393-404. https://doi.org/10.1080/17439760.2018.1460686

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly, 52*(2), 155-194. https://doi.org/10.1111/j.1533-8525.2011.01198.x

Meagher, B. R. (2022). An assessment of self and informant data for measuring intellectual humility. *Personality and Individual Differences*, *184*, 111218. https://doi.org/10.1016/j.paid.2021.111218

Meagher, B. R., Leman, J. C., Bias, J. P., Latendresse, S. J., & Rowatt, W. C. (2015). Contrasting self-report and consensus ratings of intellectual humility and arrogance. *Journal of Research in Personality*, *58*, 35-45. https://doi.org/10.1016/j.jrp.2015.07.002

Meagher, B. R., Leman, J. C., Heidenga, C. A., Ringquist, M. R., & Rowatt, W. C. (2021). Intellectual humility in conversation: Distinct behavioral indicators of self and peer ratings. *The Journal of Positive Psychology, 16*(3), 417-429. https://doi.org/10.1080/17439760.2020.1738536

National Science Board. (2014). *Science and engineering indicators*. National Science Foundation.

Nisbet, M. C., & Scheufele, D. A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany, 96*(10), 1767–1778. https://doi.org/10.3732/ajb.0900041

Palan, S., & Schitter, C. (2018). Prolific Academic: A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. *The Counseling Psychologist, 41*(4), 568-600. https://doi.org/10.1177%2F0011000012445176

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163. https://doi.org/10.1016/j.jesp.2017.01.006

Peng, C. Y. J., Harwell, M., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. *Real data analysis, 3178*.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*(6), 1774–1784. https://doi.org/10.3758/s13423-017-1242-7

Pew Research Center. (2013). *Public's Knowledge of Science and Technology*. https://www.pewresearch.org/politics/2013/04/22/publics-knowledge-of-science-and-technology/

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. Routledge.

Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J., Snow, N.E., & Jayawickreme, E. (2021). Clarifying the Content of Intellectual Humility: A Systematic Review and Integrative Framework. *Journal of Personality Assessment*, 1-13. https://doi.org/10.1080/00223891.2021.1975725

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

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177%2F0146167202286008

Saad, L. (2019, January 8). U.S. still leans conservative, but liberals keep recent gains. *Gallup*. https://news.gallup.com/poll/245813/leans-conservative-liberals-keep-recent-gains.aspx

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*, 3–15. https://doi.org/10.1177%2F096228029900800102

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177. https://doi.org/10.1037//1082-989x.7.2.147

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling psychology*, *57*(1), 1. https://doi.org/10.1037/a0018082

Simonsohn, U. (2014). [17] No-way interactions. Data Colada. http://datacolada.org/2014/03/12/17-no-way-interactions-2/

Sinatra, G. M., Kienhues, D., & Hofer, B. K. (2014). Addressing Challenges to Public Understanding of Science: Epistemic Cognition, Motivated Reasoning, and Conceptual

Change. *Educational Psychologist*, *49*(2), 123–38. https://doi.org/10.1080/00461520.2014.916216.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342. https://psycnet.apa.org/doi/10.1037/0022-0663.89.2.342

Stanovich, K. E., & Toplak, M. E. (2019). The need for intellectual diversity in psychological science: Our own studies of actively open-minded thinking as a case study. *Cognition*, *187*, 156-166. https://doi.org/10.1016/j.cognition.2019.03.006

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7th ed.). Pearson.

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*(3), 506-519. https://doi.org/10.1037/a0028857

Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, *94*(3), 509–539. https://doi.org/10.1111/phpr.12228

Zachry, C. E., Phan, L. V., Blackie, L. E., & Jayawickreme, E. (2018). Situation-based contingencies underlying wisdom-content manifestations: Examining intellectual humility in daily life. *The Journals of Gerontology: Series B*, *73*(8), 1404-1415. https://doi.org/10.1093/geronb/gby016

Zettler, I., Lang, J. W., Hülsheger, U. R., & Hilbig, B. E. (2016). Dissociating indifferent, directional, and extreme responding in personality data: Applying the three-process

model to self-and observer reports. *Journal of Personality*, *84*(4), 461-472. https://doi.org/10.1111/jopy.12172

Table 2.1

Correlations, Means, and Standard Deviations for Primary Study Variables.

Variable	1	2	3	4	5	6	7	8	9
1. GIHS	0.84[1]	-0.32**	-0.21**	0.05	-0.13**	-0.09	-0.02	0.01	0.19**
2. BAGW^{r,t}	-0.34**	0.94[1]	0.57**	-0.09	0.27**	0.25**	0.16**	0.14**	-0.14**
3. Political Orientation	-0.25**	0.56**	0.84[2]	-0.10*	0.29**	0.33**	0.33**	0.24**	-0.07
4. OSI	0.04	-0.12**	-0.12**	0.79[1]	-0.44**	-0.45**	-0.37**	-0.18**	0.14**
5. IEUB Mean Score	-0.13**	0.28**	0.30**	-0.45**	0.85[1]	0.90**	0.65**	0.48**	-0.20**
6. IEUB Endorsed	-0.08	0.24**	0.33**	-0.43**	0.91**	0.91[1]	0.76**	0.57**	-0.14**
7. IEUB Certain	-0.08	0.18**	0.31**	-0.35**	0.70**	0.77**	0.91[1]	0.56**	-0.06
8. IEUB Endorsed/Researched	-0.04	0.14**	0.23**	-0.17**	0.57**	0.62**	0.64**	0.87[1]	0.04
9. BSQ	0.17***	-0.12**	-0.01	0.11*	-0.17**	-0.11*	-0.04	0.03	0.81[1]
M	4.01	0.65	-0.37	13.04	2.53	11.60	4.45	3.50	0.91
SD	0.66	2.75	2.75	3.33	0.70	8.00	5.72	4.47	0.87

Note. GIHS = General Intellectual Humility; BAGW^{r,t} = Reversed and Log-Transformed Belief in Anthropogenic Global Warming; OSI = Ordinary Science Intelligence 2.0; IEUB = Inventory of Epistemically Unwarranted Beliefs; BSQ = Bias Blind Spot Questionnaire. Pearson correlations below diagonal; Spearman correlations above diagonal. Reliability estimates on the diagonal. [1] McDonald's Omega. [2] Cronbach's Alpha. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2.2

GIHS Correlations with IEUB Subscales.

Subscale	Mean Score	Count Endorsed	Count Endorsed as Certain
Paranormal	-0.06	-0.08	0.01
Religion	-0.21***	-0.12***	-0.11*
Health	-0.13**	-0.08	0.03
Extraordinary Life	-0.03	-0.01	0.06
Conspiracy	-0.06	-0.06	-0.04
Ghosts	-0.05	-0.01	0.01

Note. IEUB = Inventory of Epistemically Unwarranted Beliefs. Pearson correlations for Mean Scores and Spearman correlations reported for Count Endorsed and Count Endorsed as Certain. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2.3

Two-Way Interaction Model of BAGW Reverse Scored and Log-Transformed

Variable	B	SE	β	*t*
Constant	1.39	0.13		10.88***
GIHS	-0.18	0.03	-0.21	-5.59***
Political Orientation (PO)	0.10	0.04	0.53	2.66**
PO X GIHS	0.00	0.01	-0.02	-0.11

Note. BAGW = Belief in Anthropogenic Global Warming. GIHS = General Intellectual Humility Scale. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 2.4

Main Effects Model of BAGW Reverse Scored and Log-Transformed

Variable	B	SE	β	*t*
Constant	1.40	0.13		11.13***
GIHS	-0.18	0.03	-0.22	-5.69***
Political Orientation	.10	0.01	0.51	13.43***

Note. BAGW = Belief in Anthropogenic Global Warming. GIHS = General Intellectual Humility Scale. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 2.5

Models of BAGW Reverse Scored and Log-Transformed with Political Orientation, GIHS, and OSI as predictor variables

	Model 1 Main Effects				Model 2 2-way Interactions				Model 3 3-way Interactions			
Variable	B	SE	β	*t*	B	SE	β	*t*	B	SE	β	*t*
Intercept	1.50	0.15	-	10.20***	1.77	0.54	-	3.29**	1.76	0.56	-	3.18**
Political Orientation (PO)	0.10	0.01	0.50	13.20***	0.07	0.05	0.37	1.46	0.08	0.18	0.39	0.42
GIHS	-0.18	0.03	-0.22	-5.68***	-0.24	0.13	-0.30	-1.85	-0.24	0.14	-0.30	-1.79
OSI	-0.01	0.01	-0.05	-1.32	-0.03	0.04	-0.18	-0.71	-0.03	0.04	-0.18	-0.70
PO X GIHS					0.00	0.01	-0.01	-0.04	-0.00	0.04	-0.03	-0.03
OSI X PO					0.00	0.00	0.15	1.00	0.00	0.01	0.13	0.13
GIHS X OSI					0.01	0.01	0.16	0.53	0.01	0.01	0.16	0.52
GIHS X OSI X PO									0.00	0.00	0.02	0.02

Note. BAGW = Belief in Anthropogenic Global Warming; GIHS = General Intellectual Humility Scale; OSI = Ordinary Science Intelligence 2.0. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 2.6

Interaction Model of Single-Item BAGW1 Reverse Scored and Log-Transformed

Variable	B	SE	β	*t*
Constant	1.22	0.15		7.90***
GIHS	-0.12	0.04	-0.13	-3.25**
Political Orientation (PO)	0.14	0.01	0.60	2.88**
PO X GIHS	-0.01	0.01	-0.12	0.58

Note. BAGW1 = Belief in Anthropogenic Global Warming single item indicator. GIHS = General Intellectual Humility Scale. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 2.7

Main Effects Model of Single-Item BAGW1 Reverse Scored and Log-Transformed

Variable	B	SE	β	*t*
Constant	1.24	.15		8.18***
GIHS	-.13	.04	-.13	-3.39***
Political Orientation	.11	.01	.49	12.20***

Note. BAGW1 = Belief in Anthropogenic Global Warming single item indicator. GIHS = General Intellectual Humility Scale. * $p < .05$. ** $p < .01$. *** $p < .001$

Table 2.8

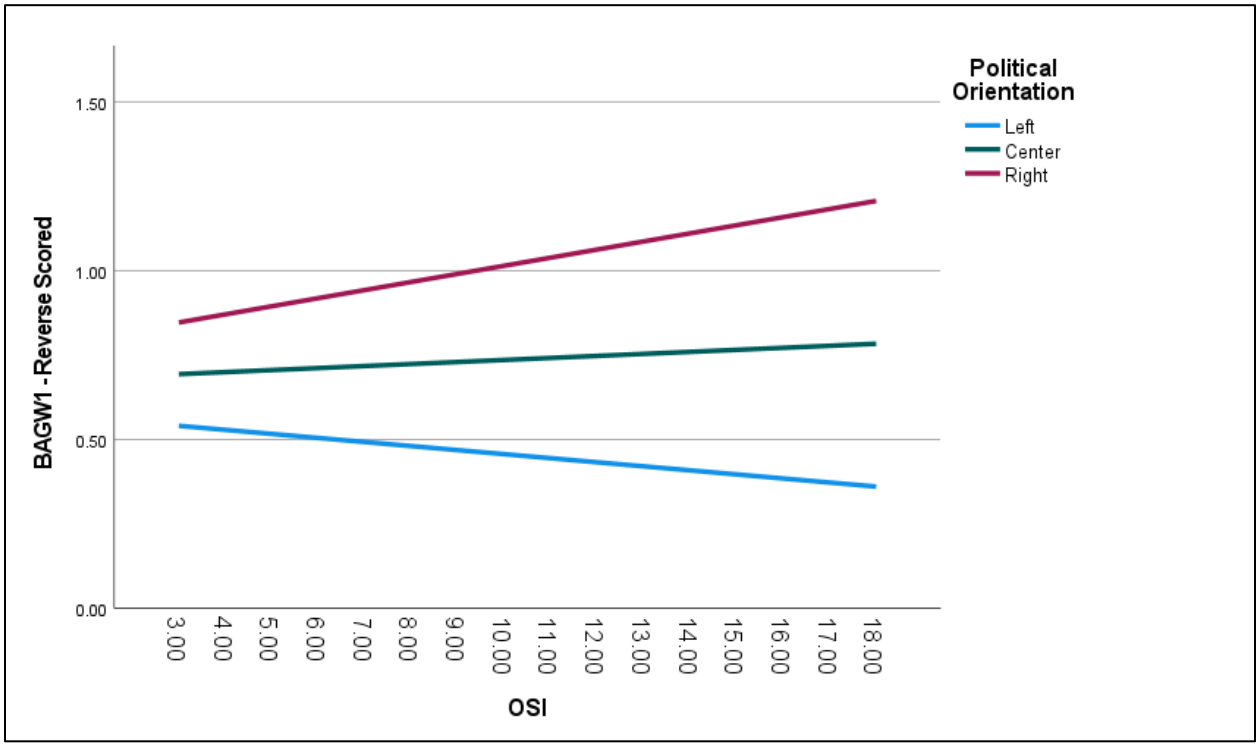
Models of Single-Item BAGW1 Reverse-Scored and Log-Transformed with Political Orientation, GIHS, and OSI as Predictor Variables

	Model 1 Main Effects				Model 2 2-way Interactions				Model 3 3-way Interaction			
Variable	B	SE	β	*t*	B	SE	β	*t*	B	SE	β	*t*
Intercept	1.18	0.18	-	6.63***	2.14	0.65	-	3.31**	2.12	0.67		3.17**
Political Orientation (PO)	0.11	0.01	0.49	12.19***	0.05	0.06	0.20	0.77	0.06	0.22	0.29	0.29
GIHS	-0.13	0.04	-0.14	-3.39**	-0.37	0.16	-0.39	-2.31*	-0.36	0.16	-0.38	-2.22*
OSI	0.01	0.01	0.03	0.67	-0.07	0.05	-0.38	-1.46	-0.07	0.05	-0.37	-1.40
PO X GIHS					-0.00	0.01	-0.06	-0.28	-0.01	0.05	-0.14	-0.15
OSI X PO					0.01	0.00	0.36	2.29*	0.01	0.02	0.27	0.27
GIHS X OSI					0.02	0.01	0.49	1.58	0.02	0.01	0.49	1.53
GIHS X OSI X PO									0.00	0.00	0.09	0.09

Note. BAGW1 = Belief in Anthropogenic Global Warming single item indicator. GIHS = General Intellectual Humility Scale. OSI = Ordinary Science Intelligence 2.0. * $p < .05$. ** $p < .01$. *** $p < .001$

Figure 2.1

Motivated Reasoning and Belief in Anthropogenic Global Warming



Note. Two-way interaction between political orientation and Ordinary Science Intelligence in predicting lesser belief in anthropogenic global warming as measured by the reversed and log-transformed Belief in Anthropogenic Global Warming single-item indicator.

APPENDICES

Appendix A: Demographics

1. What was your assigned sex at birth?
    a. Male
    b. Female
2. What is your gender?
3. What is your age?
4. What is your race?
    a. White/Caucasian
    b. Black/African-American
    c. Asian/Pacific Islander
    d. Hispanic/Latino/a
    e. Multiracial (please make sure to click all that apply)
    f. Other
5. What is your current marital status?
    a. Single
    b. Married/partnered
    c. Separated
    d. Divorced
    e. Widowed
    f. Other
6. What is or was your academic major? (If multiple or graduate school, list them all)
7. I have a relationship with God or a higher being.
    a. True
    b. False
8. What is your religious/spiritual affiliation (e.g., Methodist, Catholic, Judaism/Orthodox, Atheist, Agnostic)

Appendix B: Study Questionnaire

General Intellectual Humility Scale (Leary et al., 2017)

Below are a number of statements about personal characteristics. Please carefully read each item in the list and select the appropriate number indicating to what degree each statement best describes you. For instance, if a statement is **not at all like you**, choose 1, but if a statement is **very much like you**, choose 5.

1. I question my own opinions, positions, and viewpoints because they could be wrong.
2. I reconsider my opinions when presented with new evidence.
3. I recognize the value in opinions that are different from my own.
4. I accept that my beliefs and attitudes may be wrong.
5. In the face of conflicting evidence, I am open to changing my opinions.
6. I like finding out new information that differs from what I already think is true.

Inventory of Epistemically Unwarranted Beliefs (Dyer & Hall, 2018)

Each of the statements below describes a belief some people hold related to a given subject. For each subject, indicate whether **you personally believe the given statement to be true or false** by selecting the response that most closely matches your opinion. Use the following scale: 1 = Sure it's false to 5 = Sure it's true. [response options: 1 = Sure it's false, 2 = Uncertain, but think it's false, 3 = Really don't know, 4 = Uncertain, but think it's true, 5 = Sure it's true]

1. Age of the Earth: The Earth is about 4.5 billion years old.
2. Acupuncture: Inserting needles in the skin at specific points on the body can relieve symptoms and treat diseases.
3. Alien abduction: Extraterrestrial beings sometimes abduct human subjects for study, and then return them, often with no memory of the event.
4. Ancient astronauts: Extra-terrestrial beings helped ancient humans build pyramids and other ancient wonders.
5. Angels: There are benevolent spiritual beings who act as protectors or guides for human beings.
6. Animism: Spirits and souls exist in all things (animals, plants, inanimate objects) and can affect the material world.
7. Anti-Vaccination Movement: Routine childhood vaccines cause developmental problems such as autism in some children.
8. Area 51: The US government is hiding evidence of alien spacecraft and creatures under high security at places like Area 51 in Nevada.
9. Astrology: The location of the planets and stars at the time of one's birth are directly related to one's personality and path of life.

10. Axial Tilt: The seasons on Earth are caused by the tilt of the Earth on its axis as it orbits the Sun.
11. Bermuda Triangle: The Bermuda Triangle is a special location on the planet that, for unknown reasons, causes ships and aircraft to crash or disappear more often than anywhere else on the planet.
12. Big Bang Theory: The beginning of the universe is best explained by the Big Bang Theory.
13. Bigfoot: A large ape-like mammal lives in some northwestern US forests and is sometimes encountered by hikers.
14. Brain Capacity: Humans only use about 10% of their brains at any one point in time.
15. Coelacanth: A large fish (up to six feet long), barely changed by evolution for hundreds of millions of years, long believed to be extinct, is still living in the world's oceans.
16. Chemtrails: The white streaks left in the sky behind airplanes contain chemicals released purposely on the population for nefarious purposes.
17. Chiropractic: Manipulation of the back can effectively cure most physical maladies by aligning the spinal column.
18. Chicxulub Asteroid: Approximately 65 million years ago a large asteroid impact caused the extinction of the dinosaurs.
19. Chupacabra: A predatory animal lives unknown to biology in the southwestern US and Latin America and attacks livestock and drains their blood.
20. Climate Change: Alterations of the atmosphere due to human activity are causing a dangerous warming of the Earth.
21. Creationism: All forms of life on Earth were created at one time in history and have not changed.
22. Faith Healing: People can be healed of serious illness by a charismatic religious leader who lays on hands and/or casts out demons.
23. ESP (extra-sensory perception): It is possible to communicate mind-to-mind without using the five known senses.
24. Full Moon: A full moon causes people to behave oddly and sometimes violently.
25. Ghosts: Disembodied spirits can occasionally communicate with or be seen by living people.
26. GMO Risk (Genetically Modified Organisms): Foods produced using GMOs present a risk to human health.
27. Graphology: Characteristics of handwriting (shape and size of letters etc.) can reflect personality traits of the writer.
28. Haunting: Houses can be inhabited by the spirits of those who died in unusual ways.
29. Heaven and Hell: Human behavior on Earth will be rewarded or punished in the afterlife by condemnation to hell or reception in heaven.
30. Heliocentrism: The Sun is the center of the solar system and the Earth revolves around it.

31. Holocaust: The Nazi Holocaust of the 1930s and 1940s caused the deaths of over 11 million people, many of whom were Jewish.
32. Homeopathy: Alternative medicine using natural substances, very highly diluted, can effectively treat symptoms that would be caused by the very same substances.
33. Human Evolution: Humans and chimpanzees once shared a common ancestor, and these species diverged millions of years ago.
34. Lucky Numbers: Some numbers and dates are more lucky or unlucky than others, such as Friday the 13th or the number 7.
35. Mars Rover: NASA is currently exploring the surface of Mars with remotely controlled automated motor vehicles.
36. Moon Landing: The United States landed humans on the Moon and returned them to Earth between 1969 and 1972.
37. Mozart Effect: Listening to classical music in early childhood produces improved math skills.
38. Near Death Experiences: Some people who are close to death but are then revived, directly experience the afterlife.
39. Ouija Board: Ouija boards can be used to contact the spirit world, whose spirits answer questions by guiding the participants' hands on a device that spells out a response.
40. Prayer for Healing: Prayer is effective for treating people with illnesses and injuries.
41. Psychic Detectives: Some people can use psychic abilities to solve crimes by identifying where a body is buried or sensing whether or a kidnap victim is alive or dead.
42. Psychokinesis: Some people can move objects solely with the power of their mind.
43. Reincarnation: After death, we come back as the same spirit in a new body, occasionally with memory of past lives.
44. Satanic Ritual Abuse: Satanic cults in many communities in the US kidnap children for ritual sexual abuse.
45. Self-Esteem: Frequent praise from parents and teachers helps to raise children's self-esteem, which promotes success in life.
46. Sept 11 Conspiracy: Members of the US government were involved in the planning and execution of the attacks on the Pentagon and World Trade Center on September 11, 2001.
47. Sugar: Eating a lot of sugar makes kids hyper.
48. Therapeutic Touch: Some health professionals have the power to heal others by moving the hands over the patient's energy field to balance it, and transferring personal energy to the patient, thus healing them.
49. UFOs (Unidentified Flying Objects): Spacecraft piloted by beings not from Earth sometimes visit Earth.
50. Water Fluoridation: Fluoride is added to public water supply in order to create a more compliant population that can be easily controlled by a corrupt government.

Ordinary Science Intelligence 2.0 (Kahan, 2017)

1. All radioactivity is man-made.
    a. True
    b. False
2. Lasers work by focusing sound waves.
    a. True
    b. False
3. Electrons are smaller than atoms.
    a. True
    b. False
4. Which gas makes up most of the Earth's atmosphere?
    a. Hydrogen
    b. Nitrogen
    c. Carbon Dioxide
    d. Oxygen
5. Does the Earth go around the Sun, or does the Sun go around the Earth? [only if 'earth/around sun']: How long does it take for the Earth to go around the Sun?
    a. 1 day
    b. 1 month
    c. 1 year)
6. Antibiotics kill viruses as well as bacteria.
    a. True
    b. False
7. Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure and not give the drug to another 500 people with high blood pressure and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?
    a. The first way
    b. The second way
8. A doctor tells a couple that their genetic makeup means that they've got one in four chances of having a child with an inherited illness. Does this mean that if their first child has the illness, the next three will not?
    a. Yes
    b. No
9. Does this mean that each of the couple's children will have the same risk of suffering from the illness?
    a. Yes
    b. No
10. Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number? [open ended: 50% of or equivalent]

11. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS? [open ended: 10 or equivalent]
12. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car? [open ended: 0.1% or equivalent]
13. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease. [open ended: 20 or equivalent]
14. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000? [open ended: 100 or equivalent]
15. Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not have a tumor. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 80 of them do not have a tumor and indicates incorrectly that 10 of them do have a tumor. The table below summarizes all of this information.

	Tested positive	Tested negative	totals
Actually has tumor	9	1	10
Does not have tumor	10	80	90
Totals	19	81	100

    Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor? ___ out of ___ [open ended: 9, 19]
16. If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? ___ minutes [open ended: 5]
17. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ___ cents [open ended: 5]
18. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ___ days [open ended: 47]

Belief in Anthropogenic Global Warming (Hennes et al., 2016)

For the following items, please select the option that best reflects your personal beliefs about the subject of global warming on a scale from 1 (Definitely not; Not at all likely) to 7 (Definitely, Extremely likely).

1. Do you believe that global warming is anthropogenic (caused by human behavior)?
2. Do you believe that global warming is occurring?
3. Do you believe that global warming is a hoax?
4. How likely do you think it is that, in your lifetime, the effects of global warming will be noticeable (species extinction, glacial melting, severe weather such as hurricanes, increased temperatures?

5. How likely do you think it is that global warming is occurring?
6. How likely do you think it is that scientists will eventually discover that global warming is NOT man-made after all?
7. How likely do you think it is that global warming is a hoax?

Bias Blind Spot Questionnaire (West et al., 2012)

1. Outcome Bias: Psychologists have found that people tend to judge the quality of a decision based on how the decision worked out. That is, people sometimes forget that the quality of the decision must be judged on what was known at the time the decision was made, not how it worked out, because the outcome is not known at the time of the decision. It is a mistake to judge a decision maker's ability, after the fact, based mostly on the outcome of the decision. When people do this, it is called outcome bias.
    a. To what extent do you believe you are likely to commit outcome bias? [response options: 1 = Not at all likely 5 = Very likely]
    b. To what extent do you believe the average American is likely to commit outcome bias? [response options: 1 = Not at all like me to 5 = Very much like me]
2. Framing Effect: Psychologists have shown that people tend to evaluate statements, arguments, or policies differently depending on the choice of words. This means that people's opinions of the very same policy or decision or product can be manipulated by slight changes in wording that don't change the meaning. For example, a food item labeled "98% fat free" is judged more attractive than one labeled "contains 2% fat." When people's opinions are manipulated based on a rewording that does not change the meaning, this is termed a framing effect.
    a. To what extent do you believe you are likely to be susceptible to framing effects? [response options: 1 = Not at all like me to 5 = Very much like me]
    b. To what extent do you believe the average American is likely to be susceptible to framing effects? [response options: 1 = Not at all like me to 5 = Very much like me]
3. Base-Rate Neglect: Psychologists have shown that people tend to ignore overall probabilities when judging how likely something is and instead focus too much on the specific situation. For example, when judging the likelihood of a shark attack, people tend to focus on a news report of a single attack, rather than on the fact that although several millions of people swim in ocean water, only a few people are killed by sharks every year. When people focus on the specific example and ignore the overall probability, this is termed base-rate neglect.
    a. To what extent do you believe you are likely to commit base-rate neglect? [response options: 1 = Not at all like me to 5 = Very much like me]
    b. To what extent do you believe the average American is likely to commit base-rate neglect? [response options: 1 = Not at all like me to 5 = Very much like me]
4. Psychologists have found that people tend to rate conjunctions of events (situations where two or more events must each happen) as too likely. Conjunctions of events become less likely as the number of events grows. For example, (A) people might estimate that next year there is a 1% chance that a fire in California will kill 200 people. At the same time, (B) they might estimate that next year there is a 3% chance that an earthquake in

California will cause a fire that will kill 200 people. However, if Event B (both earthquake and fire) happens, then Event A (fire) also happens, so Event A can't be less likely. When people fail to lower the probabilities as the number of conjoined events grows, this is called a conjunction error.

   a. To what extent do you believe you are likely to be susceptible to conjunction effects? [response options: 1 = Not at all like me to 5 = Very much like me]
   b. To what extent do you believe the average American is likely to be susceptible to conjunction effects? [response options: 1 = Not at all like me to 5 = Very much like me]

5. Anchoring and Adjustment: Psychologists have found that people making numerical estimations tend to focus on any number that is available to help them. This is a good strategy, except in situations where the available numbers are unrelated to the quantity we are trying to estimate. For example, people report fewer headaches when they are asked: "How many headaches do you have a month— 0, 1, 2— how many?" than when they are asked: "How many headaches do you have a month—5, 10, 15— how many?" When our estimations are affected by quantities that are irrelevant to what we are estimating, this is called an anchoring effect.
   a. To what extent do you believe you are likely to be susceptible to anchoring effects? [response options: 1 = Not at all like me to 5 = Very much like me]
   b. To what extent do you believe the average American is likely to be susceptible to anchoring effects? [response options: 1 = Not at all like me to 5 = Very much like me]
6. Myside Bias: Psychologists have found that people do not evaluate the evidence fairly when they already have an opinion on the issue. That is, they tend to evaluate the evidence as being more favorable to their own opinion than it actually is. When people do this, it is called myside bias.
   a. To what extent do you believe you are likely to be susceptible to myside bias? [response options: 1 = Not at all like me to 5 = Very much like me]
   b. To what extent do you believe the average American is likely to be susceptible to myside bias? [response options: 1 = Not at all like me to 5 = Very much like me]
7. Cellphone Hazard: Researchers have found that drivers are four times more likely to be involved in a serious auto accident during those times when they are talking on cell phones. This effect has been called the cell phone hazard.
   a. To what extent do you believe that you are (or would be more likely to be) more hazardous during times when you drive while using a cell phone? [response options: 1 = Not at all like me to 5 = Very much like me]
   b. To what extent do you believe the average American is (or would be more likely to be) more hazardous during times when they drive while using a cell phone? [response options: 1 = Not at all like me to 5 = Very much like me]