Stable Individual Differences in Overconfidence

by

Matthew Asher Lawson

Business Administration
Duke University

Date:_____

Approved:

_____
Richard P. Larrick, Supervisor

_____
Jack B. Soll

_____
Hemant Kakkar

_____
Sandra C. Matz

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Business Administration in the Graduate School
of Duke University

2022

ABSTRACT

Stable Individual Differences in Overconfidence

by

Matthew Asher Lawson

Business Administration
Duke University

Date:_____
Approved:

_____
Richard P. Larrick, Supervisor

_____
Jack B. Soll

_____
Hemant Kakkar

_____
Sandra C. Matz

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Business Administration in the Graduate School of
Duke University

2022

Copyright by
Matthew Asher Lawson
2022

Abstract

Overconfidence has primarily been studied in one of three forms; overestimation, overplacement, or overprecision. These forms differ in their definitions, the levels of overconfidence they exhibit, and their responses to changes to task difficulty. This has led some to conclude that the different measurements of overconfidence do not share a common, dispositional component, but rather are empirically distinct. In two time-lagged studies, we show that at an individual-level, i) different elicitations of overconfidence are positively correlated, and ii) measures of overconfidence are stable over short and long time periods. We conclude from this evidence that there is a dispositional component to overconfident beliefs and that the three forms are empirically related. We also introduce a new measure of overconfidence that is less dependent on the idiosyncrasies of individual elicitations: core overconfidence. We find that core overconfidence is significantly associated with a range of individual difference measures (e.g., narcissism), and can be measured efficiently using a short three-item scale. Overall, by studying the same respondents' behaviors across elicitations, domains, and time periods, we find strong evidence for stable, individual differences in overconfidence.

Contents

Abstract iv

List of Tables viii

List of Figures ix

Acknowledgements x

Chapter 1: Stable Individual Differences in Overconfidence 1

1.1 Forms of overconfidence 1

1.2 Overview of the core arguments of the proposal 5

1.3 Evidence for situational effects and the uniqueness of elicitations 7

1.4 Evidence that different forms of overconfidence may share a dispositional component 13

1.4.1 Split-sample stability 19

1.4.2 Across domain stability 20

1.4.3 Across elicitation stability 21

1.5 Understanding of the nature of overconfidence 22

1.6 Individual difference measures and core overconfidence (correlation with traits) 27

Chapter 2: Study 1 33

2.1 Preface 33

2.2 Methods 35

2.2.1 Participants 35

2.2.2 Materials 35

2.2.3 Procedure 37

2.2.4 Coding the data 39
2.2.5 Results 41
2.2.5.1 Nature of overconfidence 41
2.2.5.2 Measuring core overconfidence 43
2.2.5.3 Association between core overconfidence and specific contexts 45
2.2.5.4 Core overconfidence and overclaiming 46
2.2.5.5 Overconfidence and individual differences 48
2.3 Second experimental session 49
2.4 Discussion 53
Chapter 3: Study 2 56
3.1 Preface 56
3.2 Methods 58
3.2.1 Participants 58
3.2.2 Materials 58
3.2.3 Procedure 61
3.2.4 Coding the data 62
3.2.5 Results 64
3.2.5.1 Nature of overconfidence 64
3.2.5.2 Estimating core overconfidence 66
3.2.5.3 Confidence, accuracy, and overconfidence 68
3.2.5.4 An alternate measure of core overconfidence 72
3.2.5.5 Stability over time 74

3.3 Discussion 78

Chapter 4: General Discussion 82

4.1 Summary and contributions 82

4.1.1 The existence of a dispositional component of overconfidence 82

4.1.2 What is the nature of this dispositional component? 84

4.1.3 Core overconfidence and individual differences 85

4.2 Limitations 86

4.3 Future directions 93

Chapter 5: Conclusions 97

Appendix A: Scales and measures 98

Appendix B: Supplemental study 115

References 132

Biography 142

List of Tables

Table 1: Example overconfidence items for each of the three forms 4

Table 2: Example individual difference measures categorized 31

Table 3: The correlations between the 12 measures of overconfidence..... 42

Table 4: Correlations among Time 1 (rows) and Time 2 (columns) measures of each variable. 50

Table 5: Correlations among Time 1 and Time 2 measures of the forms of overconfidence..... 51

Table 6: Levels of overconfidence in each elicitation. 65

Table 7: Correlations among each form of overconfidence..... 66

Table 8: Correlations among overconfidence, confidence, and accuracy measures..... 69

Table 9: Correlations among measures of overconfidence over two time periods. 75

Table B.1: The structure of the data in the supplemental study. ……118

Table B.2: Overconfidence in each domain and elicitation across Study 1 and the supplemental study. …….121

Table B.3: Individual differences and correlations with specific types of overconfidence. ……..128

List of Figures

Figure 1. A model of overconfident beliefs. 23

Figure 2. Predicted overconfidence on original response scale for different levels of core overconfidence and OCT. 91

Acknowledgements

Thank you to my committee of Rick Larrick, Jack Soll, Hemant Kakkar, and Sandra Matz. The combination of your feedback has been invaluable—both now and over the course of my time at Duke—and it is a true honor to have the chance to hear from all of you, in both the proposal and the defense. I am also very grateful to my committee for their full support in my accelerated timeline, which unfortunately also renders me relatively pressed for time in writing my acknowledgements. I could fill another 35,000 words with reflections on the kindness done unto me in the past four years (but I will spare everybody even further reading!)

To Rick, who always listened to my perspective, even before it was worthwhile to do so. Such a combination of humility, intellect, and kindness is a rare thing in the world. You taught me not only to do science and pursue the truth, but to do so in a way that respects the dignity of others, both inside and outside of the field. You are driven by a sense of moral duty in a way that is an inspiration to me, as a researcher and as a person. You are willing and able to provide support on any topic, research or otherwise. I could not feel more supported, for which I am eternally grateful.

To Jack, for your endless commitment to verifying the details. This is something that I appreciate more and more every day. When at the point of submitting a final draft, you are still probing and questioning our most basic assumptions, I can't help but smile. The rigor with which you apply yourself is a phenomenal asset to those around you and

the world at large. I thank you for taking the time for long lunches where we skirt across all manner of topics (relevant and not-so-relevant). These were some of my favorite times at Duke, and I hope they can continue (though more sporadically) in France and beyond.

To Hemant, for pulling back the curtains on research to reveal what is beneath. As you know, left to my own devices I would spend most of my time analyzing data, but you taught me how to deeply scrutinize what numbers *really* mean. What does a statistical test teach us about theory? And how does a finding reconcile with other research? In meetings, walks around campus, and watching football games, you illuminated a whole world of connections that were previously in the dark for me. I am also blown away by your endless devotion to the doctoral students at Duke—when given a choice you will always prioritize work with a student over any other task, and that's an immense gift to our doctoral community. As I move forward I am struck by remarkable insights from journal club and our other meetings that will stay with me as I navigate the world. Thank you.

To Sandra, for maintaining an utter commitment to ensuring both the well-being and progress of those around you. Working with you exposed me to an entirely new field of research, but one where I never felt lost given your unerring guidance. Your help when I was stuck in DC and invitation to NY when I was stuck in the mud were both profound moments that also served as lessons for how to treat others. It is not only

invigorating but also wholeheartedly *fun* to work with you and I can't wait to resume everything that we have in the pipeline. I am so glad we got to catch up more at SPSP and look forward to conferences for many years to come!

To the rest of Duke faculty and my co-authors, for the support, small and large, but without which nothing would have been possible. They say it takes a village to raise a child and I feel that the same stands for a PhD student. You of course all know that I think this because I emailed everyone the same thought after a particularly enjoyable departmental event. Our seminars and social events have been highlights of my time at Duke and I am always one of the last to leave. It is a great shame that worldly circumstances reduced the number of events we had to get to know each other, but regardless I am so grateful for the time that we did have.

To the PhD students at Duke, I share similar sentiments. For one, I was blessed with a cohort (Dani and Jess) where we did nothing but support each other—professionally, emotionally, and in any other way that we could. This defined my experience at Duke, making any challenge feel surmountable and opening unbounded doors of possibility. I am sincerely grateful to Dani and Jess for their support in my hardest times at Duke, as well as for their professional help. Dani always provides a balancing perspective when I am making histrionic claims, and Jess helps me to see more charitable perspectives about my own work. More broadly, I am grateful to the doctoral community for all of the help and fun we had along the way. The older

students were always available for any advice and support we needed, and the friends I have made within the Fuqua community will be lifelong. For this I am grateful.

Beyond the faculty and students, I am thankful for a plethora of other people who made my time at Duke smooth, enjoyable, and possible. I thank Duke's administration staff, for their commitment and support with all of the practical side of executing research and travel. I thank them for their patience as I submitted wrong and outdated forms, missed deadlines, and made any number of errors that they could have harshly penalized, but chose not to. I thank all of the other support staff—from IT, payroll, and many other departments—without which I couldn't have taught or researched how I wanted to. I thank the staff at Duke for providing us with a comfortable and welcoming place to work. For the cafeteria workers with warm smiles in the mornings, and the chats with the cleaners in the mornings (who are frequently the only people there when I like to arrive!). In particular, I thank our Lovely housekeeper, who has become a shining light in my mornings since she arrived and whose endless kindness is a true treasure.

I thank the friends and animals who have supported me along the way. Doing a PhD isn't about doing one good day's work. It's about being consistent across a long period. That means working through times where your personal life is hard, or when you don't want to. As you try to pick yourself up and work through life's obstacles, the friends around you and resources you have determine how easy that is or whether it's

even possible. I am blessed to have incredible friends—both in the US and Europe—who never faltered in being there for me. There are too many to name, but you know who you are. I acknowledge that I never would have got to the stage of writing a PhD without those who were there for me. Thank you.

Finally, I thank my loving partner and wonderful family for the support now and over the course of my time in the US. Love and family are two of the most important things in life, and frequently offer me perspective when I have become lost in the weeds. I thank my partner for listening to any research idea that I have and providing engaging and thoughtful questions (and validation when it is needed). I thank my father for the time spent studying mathematics together (and everything else), and my mother for teaching me that kindness need not be complex. I thank my grandparents for teaching me equanimity and showing that love can endure, my brother for his selfless care, and my extended family for the love and support with which they have touched my life. Thank you for your support now and as I move forward in my career.

Chapter 1: Stable Individual Differences in Overconfidence

Overconfidence is a widely studied phenomenon across academic disciplines ranging from psychology (Oskamp, 1965), to economics (Galasso & Simcoe, 2011), to strategy (Camerer & Lovallo, 1999), accounting (Malmendier & Tate, 2005a), and law (Goodman-Delahunty et al., 2010). It has likely captivated such interest due to the potentially devastating *consequences* of people's overconfident beliefs, rather than the nature of the beliefs themselves. Egregious anecdotes, such as the risks taken by pre-2008 traders, or the feeling of impunity displayed in corporate fraud cases such as Enron, are vivid in the mind, and highlight the importance of this work. However, such broad application of the concept of overconfidence can also inhibit understanding it, as the operationalizations of overconfidence are so wide ranging that it can be hard to work out exactly what we mean when we say *overconfidence*.

1.1 Forms of overconfidence

One organizing framework that is used to describe different forms of overconfidence is Don Moore and colleagues' categorization of overconfidence as one of three forms; overestimation, overplacement, or overprecision (Moore & Healy, 2008). These different forms of overconfidence differ how they are defined and elicited. We illustrate each next.

Overestimation is described as overestimating "one's actual ability, performance, level of control, or chance of success" (Moore & Healy, 2008, p. 502). The simplest example of this is asking someone how many questions they answered correctly in a quiz. If they respond with a number that exceeds the actual answer, they are overestimating their performance. This kind of overconfidence is perhaps the most intuitive and aligned with lived experience: if you felt confident coming out of a math test but later found out you got only half right, you exhibited overestimation. In general, overestimation reflects an inflated view of one's own abilities or

performance. For a concrete example of overestimation (and the other forms of overconfidence referenced in the below paragraphs), see Table 1.

The second form of overconfidence is called *overplacement*, which specifically refers to people's evaluations of their abilities and performance *relative* to others. This form of overconfidence, which is sometimes studied under the name of "better-than-average-effects", occurs when people believe themselves to be higher up in an outcome distribution than they actually are. For example, if someone tells you they are in the top 10% of all motorcar drivers, but are in fact only a better driver than 70% of the population, this is overplacement of their driving ability. Overplacement is measured by surveying someone's beliefs about their performance relative to others, and then comparing that judgment to their actual relative position (Larrick et al., 2007): Both perceptions of one's own performance and perceptions of the performance of others could contribute to a distorted view of one's relative standing here.

Finally, *overprecision* is a form of overconfidence where people assign too much probability to the outcomes that they view most likely—thus underestimating the uncertainty in what could happen, and exaggerating the precision of their own knowledge. This occurs if people provide overly narrow confidence intervals when asked to provide a range of values about which they're X% sure that the true value will lie inside the range (Soll & Klayman, 2004). For example, one might pose the following question: *What do you think is the population of Poland*? Instead of a single response, participants would be asked to provide numeric estimates that represent different quantiles of a subjective distribution of possible answers. For example; i) A value that you're 90% sure the population is greater than and ii) A value that you're 90% sure the population is smaller than. This is one way of eliciting an interval in which a judge is 80% confident that the true value will fall. If people's 80% confidence intervals contain the true values less than 80% of the time, they are displaying overprecision. We observe such interval estimates

in life when people express margins of error (e.g., “I’ll be there in 15 minutes, give or take”; see Klayman et al. (1999) and Yaniv and Foster (1995)). Interval questions (which measure overprecision) show overconfidence to the extent that people’s 80% intervals capture the true value less than 80% of the time, suggesting that people do not allocate sufficient likelihood to outcomes that are further away from what they expect will happen.

Overprecision is the least intuitive of the three forms of overconfidence, so requires a little more explanation. Though 90% subjective confidence intervals should contain the true value 90% of the time, the actual hit rate is closer to 50% (Moore, 2020). As it turns out, people’s interval widths are relatively insensitive to the level of confidence you ask for (Teigen & Jorgensen, 2005; Yaniv & Foster, 1997). This is especially harmful when a high degree of precision is required—e.g., when forecasting crude oil prices to evaluate the merits of building a new oil refinery (Russo & Schoemaker, 1992). Overprecision also displays a lack of awareness of what one does not know—people indicating narrow intervals fail to account for things that they do not expect (Macenczak et al., 2016).

Scholars have gone further to include other self-enhancements as forms of overconfidence (e.g., the illusion of control, per Moore and Healy (2008)), but for the present research we limit our scope to those that include specific measurements of performance, confidence, and their deviations. Below, we provide concrete examples of each of these three forms of overconfidence.

Imagine being asked to take a geography quiz that asks about the population sizes of countries. The first row shows the kind of knowledge tested. The second row then shows how one’s confidence in performance can be measured. The third row then shows how an accuracy standard can be calculated, allowing a comparison between stated confidence and a person’s actual record of being right. The difference between the two is *overconfidence*.

Table 1: Example overconfidence items for each of the three forms

	Overestimation	Overplacement	Overprecision
Sample Test Item	Which of these countries has a larger population, Sri Lanka or Singapore?	Which of these countries has a larger population, Sri Lanka or Singapore?	Estimate the population of Sri Lanka.
Confidence measure	*Two-answer-questions (2AQ)* What do you think is the chance of your answer being correct? (50-100%)	*Comparison to others* Compared to other participants who are taking this survey, what percentile do you think you fall in, in terms of how accurate your answers were?	*80% interval* We're going to ask you for two things: 1) A value that you're 90% sure the quantity is greater than 2) A value that you're 90% sure the quantity is smaller than
How overconfidence is calculated across multiple items	Take the average of the confidence measures. Subtract the percentage of questions answered correctly.	Rank the respondents based on performance across the items. Subtract their true performance percentile from their estimated performance percentile.	In the above example, 80% confidence intervals were elicited. Take the percentage of intervals that included the true values. Subtract 80%.

Note. Some research into overplacement compares participants' average estimates of their own performance to what the average estimate *should* be (50%). In our typology, this only captures the *confidence* aspect of overplacement at the individual level, rather than the overconfidence.

We should note that there is a long history of statistical critiques of overconfidence—with some researchers questioning whether overconfidence even constitutes a reasoning error (Gigerenzer et al., 1991), and others showing how noise and other statistical artefacts can contribute to the observed patterns of results (Klayman et al., 1999; Erev et al., 1994). We do not directly review these literatures but instead raise their insights as relevant when we discuss why different forms of overconfidence might appear distinct (Section III) or not (Section IV).

1.2 Overview of the core arguments of the proposal

The division of overconfidence into the three forms listed above has generated a wealth of research into each form's behavior and practical relevance. They differ in important ways—both in the levels of overconfidence they exhibit, their relationships with each other, and their relationships with external variables. Researchers have asked questions such as, do people display more overconfidence in one elicitation than another? Does someone exhibiting overprecision (i.e., their 80% confidence intervals contain the true value less than 80% of the time) mean they will also exhibit overestimation (i.e., will they think they got more questions right in a test than they did)? These questions primarily concern task-level variables (e.g., the average level of overconfidence in an elicitation) and level-based hypotheses (e.g., for which elicitations do people tend to display overconfidence or underconfidence?). In the present research, we ask a different set of questions. We study the behavior of the same *individuals* across tasks and time periods. This allows us to focus on a different set of insights regarding the possibility of stable associations between forms of overconfidence across measures.

Extant research has made two claims: i) individual differences in overconfidence are overwhelmed by situational factors (Moore & Dev, 2018), and ii) the three forms of overconfidence are not different manifestations of the same underlying construct, but rather empirically distinct (Moore & Healy, 2008). By changing our focus to analyzing the consistency with which individuals display *relatively higher* overconfidence across different tasks, we have reached different conclusions. The core claims of the present research are i) that there is a dispositional component at the core of overconfidence that is partially captured by different elicitations and ii) as a result, the three forms of overconfidence are empirically related.

We aim to show support for these two claims in three ways. First, we show that different forms of overconfidence are correlated at an individual-level. This directly tests our second claim.

Second, we extract a core measure of overconfidence from measures of different elicitations that predicts other forms of overconfidence in out-of-sample tests. This latter approach attempts to measure overconfidence in a way that is less dependent on the idiosyncrasies of individual elicitations to measure its core component. Finally, we test the stability of this core overconfidence over time to test whether it contains a dispositional component. It could be the case that situational factors affect all three forms of overconfidence (e.g., momentary variation in individual positive affect leading to elevated levels of overconfidence in each measure), which could lead to the forms being correlated at a given moment without a dispositional component of overconfidence. Testing stability over time mitigates this concern.

From these three forms of evidence, we conclude that there is a dispositional component to overconfidence, and that the different forms are meaningfully empirically related. We further propose that the dispositional component of overconfidence will be associated with previously established individual differences—both in people's self-views and cognitive styles—and that it may be identified with a short survey.

We start by reviewing the impact of situational effects and the uniqueness of the different elicitations, focusing on results that support the idea that they will not be related (Section III). These include the different levels of overconfidence obtained in *different elicitations*, as well as the role of *task difficulty* as a determinant of overconfidence, and cases where different elicitations have different relationships with *outside factors*. From there, we move to discussing the theoretical reasons why different forms of overconfidence might be related, and the existing evidence for a dispositional component of core overconfidence (Section IV). We next describe our understanding of the nature of overconfidence (Section V), before weighing the evidence that the core component of overconfidence will be associated with individual difference measures in

Section VI, as we propose to test how self-view and cognitive style measures predict core overconfidence.

1.3 Evidence for situational effects and the uniqueness of elicitations

In the following section, we review evidence that suggests that there is *no dispositional component* to overconfidence, and that the three forms of overconfidence are *not related*. These two points are related—the existence of a dispositional component underpinning each elicitation would normally imply some association—and so we group the contradictory evidence together. We primarily ask the question, what is the evidence that the three forms are not related?

There are several reasons why scholars have concluded that the three forms of overconfidence discussed so far are distinct. The first reason is their conceptual differences. Overestimation captures people inflating their own performance or attributes—their estimates are compared to the ground truth, and they frequently hold overly positive beliefs about themselves. These measures generally explicitly ask for confidence (e.g., what's your confidence in this answer, or what is your estimated performance), and are tightly tied to the self. Overplacement differs from overestimation that it measures people's perceptions of the self relative to a reference group. People are asked about their relative standing. If different reference groups are activated at different times, this could produce distinct results. Overprecision is more distinct still, in that it does not explicitly ask for confidence, but rather elicits two or more estimates that define a range of what people think are likely outcomes. This captures people's perceptions of uncertainty—in their knowledge and in the world—with people often underestimating uncertainty in both. This

form is less tied to people's self-views, and more to how they perceive uncertainty.[1] In sum, there are different elements that contribute to holding each form of overconfident belief, which could support the idea that they will not be related.

Another piece of evidence that could suggest that these forms are unrelated is the observed levels of overconfidence seen in each paradigm. For example, respondents tend to exhibit significantly more overconfidence in providing interval estimates (overprecision) than they do when answering two-answer-questions (2AQ) that elicit confidence levels and compare them to accuracy rates (overestimation; see Klayman et al. (1999); Moore & Healy (2008); Juslin et al. (1999)). Specifically, in the overprecision paradigm, people's 90% confidence intervals often capture the truth around 50% of the time, indicating a gap between confidence and accuracy of 40% (Moore, 2020). In contrast, measuring overestimation using the 2AQ paradigm yields much smaller gaps, such as 5% (e.g., an average confidence of 70% and accuracy rate of 65%). In fact, it is common to find domains in which people are not overconfident, but underconfident. Overall, people sometimes exhibit underconfidence in overestimation, but nearly universally exhibit overprecision (Budescu & Du, 2007; Winman et al., 2004; Juslin et al., 1999) with large gaps between confidence and accuracy in interval based paradigms (Klayman et al., 1999; Russo & Schoemaker, 1992). As a result, someone who displays overprecision may not also indicate positive scores in overestimation and overplacement measures, even if they were positively correlated. The fact that overconfidence varies so much across the different elicitations could suggest that situational factors dominate over any influence of a dispositional component.

This difference in the observed degree of overconfidence across elicitations has inspired research into factors that could explain it. Consider how each form of overconfidence is elicited.

---

[1] Though even here, it is not unequivocal that people always recognize that a narrower range implies less uncertainty than a wider range when they are forming their estimates.

In many overestimation paradigms, people choose between two answers, and a contrasting answer is always present (as shown in Table 1). Yet when forming interval estimates, people lack listed alternatives. This could leave people's judgments more prone to confirmation bias (Klayman et al., 1999), which would explain why a greater degree of overprecision is exhibited than other forms of overconfidence. Researchers have also focused on how overprecision differs from the other forms in that it relies on a process of *adjustment*. People may form a central anchor as an estimate and do not sufficiently adjust from that estimate when forming an interval estimate (Tversky & Kahneman, 1974), yielding the high degrees of overprecision observed. This seems to imply that making a central estimate of an interval more salient would increase overconfidence, yet Soll and Klayman (2004) find the *least* overconfidence in interval estimates when explicitly requiring a central estimate (see also Juslin et al., (1999); Block and Harper (1991)), so further explanation is required.

Much research has examined the levels of overconfidence displayed in different tasks (i.e., the size of the gap between confidence judgments and objective reality). These levels are important, as larger gaps could lead to worse consequences—though one prominent critique of overconfidence findings is that they exaggerate the degree of actual overconfidence due to non-random sampling of stimuli. Suppose a set of questions contains a number of items where judges have information that is relevant to the question, but leads them to an incorrect answer because valid cues are misleading for that specific instance (i.e., what Klayman et al. (1999) call a "contrary question"). For example, if one asked which is further west, Los Angeles, CA or Reno, NV, it is tempting to say Los Angeles. The reason is that two cues point to this answer: 1) LA is on the Pacific Ocean and Reno is landlocked; 2) The state of California is west of the state of Nevada. Both generally valid cues, however, do not take into account that southern California "bends" eastward substantially and that Reno is on the far western edge of Nevada. Reno is in

fact further west than LA. Here, people may reasonably be both wrong and confident because the cues they are using typically work: for example, Fresno *is* west of Las Vegas. Due to the effect of cue validity on overconfidence, it is important to sample questions with cue validities that are in line with the population's average—if contrary questions are oversampled, this may inflate the observed level of overconfidence. This concern can be addressed with representative stimulus sampling (Gigerenzer et al., 1991; Juslin, 1994). In short, the degree of observed overconfidence can be hard to interpret because it also depends on what items are sampled—a nuisance factor that does not tell us about the underlying psychology of the individuals answering the items.

A related critique of specifically overestimation paradigms is that there are disparities between the degree of confidence observed if you ask about a set of questions versus the average of that set's constituent items. Sometimes large discrepancies are observed between people's item-level confidence estimates and their guesses for how many answers they got right overall (Gigerenzer et al., 1991). Gigerenzer et al. (1991) argue that people are cognitively better equipped to encode frequency information than to reason from probabilities. Research has shown that people's ability to estimate the number they got right is sometimes more accurate than their average item-level confidence estimates. Kahneman and Tversky's (1996) model of the "inside" versus "outside" view offers one explanation for this phenomenon, which Klayman et al. (1999) summarize as people's inability to learn what a *collection* of information implies about the chance of getting a *specific* question correct. This discrepancy shows that there are unique determinants of the levels of overconfidence even *within* a form (overestimation), further emphasizing the important role of situational factors, which could imply that the forms of overconfidence will be unrelated. However, the strongest evidence for the uniqueness of different forms of overconfidence comes from considering their relationship with task difficulty. We summarize a version of this argument next.

One observation that might support viewing overestimation and overplacement as *related* would be them both increasing or decreasing as task difficulty increases. The converse—difficulty differentially affecting the two—could support the idea that these are distinct psychological phenomena. In fact, greater task difficulty (as measured by lower objective scores on a task) appears to increase overestimation (Erev et al., 1994), but produce underplacement (Burson et al., 2006; Larrick et al., 2007; Kruger, 1999; Moore & Small, 2007) at a task level. So, if an experimenter varied overall task difficulty, the average levels of overestimation and overplacement may move in different directions. This could suggest that these forms of overconfidence have different psychological bases. Yet others find that the difficulty effect on overestimation is driven by failing to account for the statistical artefacts of random noise in the number of "contrary questions"—questions whose answers run contrary to the information available to respondents—and disappear when such effects are accounted for (Klayman et al., 1999). This is when considering the difficulty effect at a task level (i.e., comparing difficult tasks to easier tasks). It can also be considered within tasks, across people (i.e., how are within-task accuracy and overconfidence related). Each of these approaches could lead to different conclusions.

Consider the effect of task difficulty *within* a task. One school of explanations for this observed pattern is based on the following "differential regressiveness" model, which considers the effect of task difficulty on individuals' overconfidence (Moore & Healy, 2008; Krueger & Heck, 2020): Suppose you have three variables; people's estimated performance (S), their true score (T), and their estimate of the population average score (O). Assume that people's confidence has some validity as a measure of their performance, but this relationship is regressive, and their estimates of others' performance are even more regressive ($r_{ST} > r_{OT}$, both positive; Krueger, Freestone, and MacInnis (2013); Moore and Healy (2008)). Assume also that

people believe in some self-other similarity (i.e., r_{OS} is positive). If you simulate sequences of these variables, you can then order them by the true score T as a measure of task difficulty. For those with high levels of performance (T)—and thus low task difficulty—underestimation and overplacement is observed, and vice versa for low levels of T. In combination with the task-level effects, this could also suggest that overestimation and overplacement are distinct phenomena.[2]

Finally, another piece of evidence that the different forms of overconfidence are unrelated is that there are cases where one form (e.g., overestimation) is related to another variable but another is not (e.g., overprecision). For example, Cain, Moore and Haran (2013) studied market entry in an experimental design that distinguished between different forms of overconfidence. They found that overplacement contributed most to the choice to enter a competition, whereas overestimation is likely unrelated. Kang and Kim (2021) tested the relationship between expertise, feedback, and overprecision, finding that experts indicate greater overprecision relative to novices when they receive disconfirming feedback. Though the effect of expertise on overall levels of overestimation and overplacement was reported and statistically significant, whether this relationship held for the other forms was not indicated. In other words, there are cases where some forms of oveconfidence behave in a certain way in relation to outside variables and others do not. This evidence suggests that the forms of overconfidence are distinct and do not arise from a common, dispositional overconfidence.

---

[2] The fact that the within-task model described above defines task difficulty as it is observed in the present sample leads to selecting on people's performance. This could introduce reversion effects as an artefact of how the data is sorted. For example, if you sort the data based on performance within a task, the people who do the best may have positive residuals in their performance (i.e., they do better than they "should" do). If these people's confidence is the same level that it is when they have similar experiences with a task, then they will be underconfident in that task. So, if an easy task is defined for an individual as one where they get more questions right in that task, there will be underestimation for even generally calibrated individuals due to selecting on positive residuals.

Overall, the literature proposes several reasons why the forms of overconfidence may be distinct; i) they have different definitions, ii) more overconfidence is observed in interval estimates than other forms, iii) each form has unique psychological determinants, iv) task difficulty appears to differentially affect overestimation and overplacement, and v) the different forms can have different relationships with other variables. We next consider the possibility that—despite these factors—the three forms can be partially explained by a common, dispositional factor; and will thus be empirically related at an individual level.

1.4 Evidence that different forms of overconfidence may share a dispositional component

In addition to the evidence discussed so far, there is additional work that is relevant in distinguishing whether there is likely to be a dispositional component of overconfidence. We weigh this evidence next, before summarizing our overall understanding of the nature of overconfidence in Section V. We start by discussing our approach, before considering how the antecedents of overconfidence effects inform our theorizing, and finally weighing the extant evidence for individual differences in overconfidence.

In the present research, we adopt *individuals* as the unit of analysis, asking whether someone who is more overconfident in one task is likely to be more overconfident in another task. Many approaches have focused on properties of the task (e.g., the elicitation, task difficulty) rather than the individuals responding to the task.[3] Yet we feel that ours is also a practically

[3] This has contributed to some methodological issues—e.g., the fact that many overplacement studies compare people's *average* performance estimates to 50%, rather than each person's estimated relative performance to their actual relative performance. This impedes studying individual differences in

useful question to ask: For example, if knowing that a decision maker was highly overconfident in one domain was helpful in predicting their overconfidence in another domain, this could be used to manage risk and allocate projects accordingly. This observation that people may be overconfident across situations is aligned with lay perceptions that some people—politicians and celebrities may come to mind—are more overconfident than others (Moore & Dev, 2018).[4] We argue that, based on plausible psychological antecedents of overconfidence effects, there is a theoretical basis for the proposal that different forms of overconfidence may be related at the individual level.

There are two main schools of explanations for the antecedents of overconfidence—statistical artefacts and psychological biases. We have made reference to some of the former explanations, with further work suggesting that imperfections in people's evaluations of the available information (Erev et al., 1994) and in how people map subjective confidence to a response scale (Erev et al., 1994; Ferrell, 1994) are also potential sources of overconfidence. Overconfidence is sometimes attenuated when controlling for these artefacts. Yet importantly, studies controlling for many of the ecological explanations still find overconfidence (Klayman et al., 1999; Soll & Klayman, 2004), which suggests that these cannot be the *sole* explanations of overconfidence. In the following paragraphs, we discuss why some of the psychological antecedents of overconfidence suggest that the forms of overconfidence may be related. We first consider why we might expect someone to be consistently overconfident *within* an elicitation,

---

overplacement, as the individual level measure—people's performance estimates—measures confidence rather than overconfidence.

[4] Note that without displaying consistency over time (i.e., test-retest reliability), one cannot rule out the possibility that any consistency observed across domains and elicitations could be explained by situational factors (e.g., positive affect).

before considering why someone might display a consistent degree of overconfidence *across* elicitations.

Consider overprecision, the form of overconfidence where people indicate overly narrow confidence intervals when estimating quantities. Researchers have proposed that when providing interval estimates, people first make a judgment (i.e., a point estimate) and then adjust this to reflect uncertainty, but do not adjust sufficiently (e.g., Koriat et al., 1980; Hoch, 1985; Block & Harper, 1991). This could be due to biases in the way people search for and evaluate information. For example, after forming a judgment, people's associative memories will be biased toward retrieving information consistent with that judgment, while also interpreting any contradictory information through the lens of the judgment (Klayman et al., 1999). This process will skew people's search processes towards producing unwarranted confidence in their initial judgment, though the judges may believe that their processes were unbiased (Klayman, 1995). People who engage in less reflective thinking might be generally more prone to committing such an error—whereas those who reflect more broadly may consider more information and widen their intervals appropriately. Based on this psychological process, we would expect people to consistently return relatively narrower or wider confidence intervals. In other words, if biased search processes contribute to overprecision, we would expect some consistency in the extent to which people display this tendency.

Now consider overestimation (inflated perceptions of one's own abilities or performance). One explanation for these effects focuses on the heuristics and biases people use when forming judgments. Kahneman and Tversky (1996) aimed at reconciling the fact that people indicate different degrees of confidence when asked for estimates after each item versus when asked how many items they answered correctly by considering the role of heuristics in judgment. They propose that when considering a single case, people take an "inside view" and

underweight relevant statistical data in these cases, thus overstating their confidence. This allows them to maintain a high degree of confidence in the validity of specific answers, despite knowing their overall hit rate is low when they take an "outside view". This is called the "illusion of validity" (Kahneman & Tversky, 1973), and the theory was formalized in the strength-weight model (Griffin & Tversky, 1992). We argue that those who engage in less reflection might be more likely to engage in heuristic thinking, and thus overestimate their own performance. We would again expect people to *consistently* exhibit this tendency across different questions in an overestimation paradigm. People who rely on heuristic judgment to form estimates that result in overconfidence are likely to do so consistently. Now, consider these two explanations together.

These two theories of overconfidence effects above both hinge on how people process information—with biased search leading to insufficient adjustment from an anchor for overprecision, and use of heuristics leading to neglect of statistical information for overestimation. When we consider overconfidence at the level of the individual, it seems plausible that someone who engages in one type of biased search may also be more likely to engage in another. Here we suggested that both forms of overconfidence could result from intuitive thinking, which implies that those who are less dispositionally reflective may be more likely to be overconfident across elicitations. This is just one possible pathway that would predict stable individual differences in overconfidence across elicitations (i.e., a dispositional component of overconfidence). The pioneering work of Stanovich and West (e.g., 1998) has found significant evidence of individual differences in reasoning processes—supporting the idea that people may engage in biased search processes to a greater or lesser degree, which could drive individual differences in overconfidence across tasks. If this were the case, we would expect individuals to show some consistency in the relative degree of overconfidence they exhibit across different tasks. Though situational factors are undoubtedly also very important and there are

unique effects of different elicitations, we believe that there is a theory-driven case for expecting a dispositional *component* of overconfidence. Having established the theoretical basis for our hypothesis, we next turn to considering the extant evidence for such a core overconfidence factor.

To do so, it is first necessary to distinguish between different ways in which overconfidence could be observed as a stable tendency observed across individuals. We outline four different types of evidence below.[5] Consider a study participant who answers 10 questions on the populations of countries, each time picking which of two countries was more populous (e.g., "Of India and Belgium, which country has the larger population?"). After each choice, they state their confidence that their answer is correct. If there was no dispositional component to overconfidence, we might expect that people's overconfidence in the first 5 questions would be unrelated to their overconfidence in the last 5 questions. After all, they are different questions. Alternatively, it might be the case that the two measures of overconfidence (calculated over the first five and last five questions, respectively) would significantly correlate (since, for example, the person's actual abilities or self-views in a domain could be stable). This is one measure of individual-level stability in overconfidence that we call *split-sample*. This type of consistency occurs within a domain, so could be partially explained by domain-specific knowledge.

Now suppose the same person from the previous example also completes a task where they answer 10 questions regarding the maximum temperature in July in U.S. cities. After answering each question, they state their confidence that their choice is correct. Another approach would calculate the respondent's overconfidence in the population domain and the temperature

---

[5] There is a fifth type of consistency—stability over time—which we do not expand upon, as we are not currently aware of any evidence testing this stability (except the present research). In a related study, Bruine de Bruin et al. (2007) establish the test-retest reliability of a measure of *calibration* which uses 34 overestimation questions, equally penalizing over and underconfidence as deviations from perfect calibration. They find a correlation of $r = .47$ ($p < 001$) between the two measures of calibration, with an average of 17.6 days between the sessions. Calibration is theoretically and empirically different from overconfidence, though of course related.

domain, and correlate the two. A significant correlation across domains could also provide evidence for individual-level stability. We call this measure, *across domain*. Both the population and temperature questions asked the person to provide answers and state their confidence that they were correct—all that changed was the *domain* from which the questions were drawn (country populations vs. city temperatures). So far, split-sample and across domain forms of stability both offer possible evidence for a dispositional component of overconfidence—people consistently exhibiting overconfident beliefs in a particular response format—but it is difficult to distinguish between situational and dispositional effects when considering a single type of task.

A different approach to capturing individual-level stability would ask people questions *elicited* in different ways. So far, we have proposed asking our hypothetical study participant items that used an overestimation elicitation—the respondent would provide item-level confidence estimates, and their accuracy was the number of questions answered correctly. Another method of measuring overconfidence would be using an overprecision paradigm. Here, we would instead ask the study participant to provide intervals in which they were 80% sure the true populations of countries lay. We could compute overprecision by comparing the percentage of their confidence intervals contained the true country populations to the expected rate (80%). Suppose we asked a group of people the first task we described—10 questions regarding country populations—as well as this interval-based task. Correlating the overestimation displayed in the former with the overprecision displayed in the latter would give us a measure of the stability of overconfidence *across elicitation*. This would show whether overconfidence was correlated even when varying situational features of the task. Surprisingly, this comparison is found more rarely in the literature. This comparison also directly tests our second proposition—that the forms of overconfidence will be empirically related.

The final form of individual-level stability in overconfidence we propose is the following: compute someone's overconfidence in any task, and correlate that overconfidence with established individual difference measures that measure personality traits or cognitive styles. If there is stability in overconfidence, we might expect that it would significantly correlate with established individual differences that are stable properties of individuals. Concretely, we could take the overestimation we measured in the first task (in the domain of country populations) and correlate it with a measure like the Big Five personality dimensions (Gosling, Rentfrow, & Swann, 2003). We call this form of individual stability in overconfidence *correlation with traits*.

We consider the evidence provided by the first three types of measurement that a common factor contributes to overconfidence effects in the following section, before expanding more on *correlation with traits* in Section VI. We argue that this common factor is likely to contain a dispositional component, but without verifying its stability over time it could be that this factor merely reflects situational factors affecting different measures of overconfidence in the same experimental session.

1.4.1 Split-sample stability

First, we consider evidence for a dispositional component of overconfidence based on split samples within a measure. Klayman et al. (1999) found with an overestimation paradigm that when collapsing across domains and splitting their sample into two subsamples, the observed overconfidence in these two subsamples was correlated at $r = .66$. This replicated earlier results found by Soll (1996) and Budescu, Wallsten, and Au (1997). West and Stanovich (1997) also found an internal consistency between split samples in an overestimation task of .71. In an overprecision task, the correlation between a participant's accuracy (which is collinear with overconfidence) across split-samples was even higher—$r = .84$ (Klayman et al., 1999). Glaser,

Langer, and Weber (2013) also found individual differences in split samples of interval estimates (see next section for more details).

Split-sample stability shows that within a task, people's overconfidence in different subtasks is highly correlated. This suggests that overconfidence is not *solely* generated at a question level, and at least has common elements across items within a task.

1.4.2 Across domain stability

Klayman et al. (1999) also found that overestimation was correlated within-elicitation across domains—three pairwise correlations among domains were $r = .60$, $r = .58$, and $r = .51$—the second type of evidence that there may be a dispositional component to overconfidence. West and Stanovich (1997) found similar evidence that overconfidence correlated across two tasks (a motor task that involved predicting one's performance, and a knowledge calibration task that used an overestimation paradigm), though only after a practice block of the motor task ($r = .24$, $p < .01$). Moving away from overestimation, Glaser, Langer, and Weber (2013) divided an overprecision task's results into subsets of questions on general knowledge and economics and finance knowledge questions, finding strong and significant correlations between the overconfidence observed in each (*r*s between 0.25 and 0.48, all significant). This constitutes both split-sample and across domain evidence.

There are two notable investigations that explicitly consider a core confidence factor—focusing primarily on how *confidence* judgments correlate within elicitation across domains. We discuss each next. Pallier et al. (2002) asked a large (N = 520) sample of Air Force recruits a long series of competency based questions, eliciting confidence in an overestimation format. The range of confidence elicited was 20% to 100% for 5-option multiple choice questions, and 0 to 100% for open-ended questions. They found that the correlations in *confidence* levels across tests were all significant at an alpha of .01 (with *r*s ranging from .21 to .72). They then used confirmatory

factor analysis (CFA) to test for individual differences in confidence, finding evidence of a metacognitive trait that determines the calibration of people's confidence, which was distinct from both intelligence and personality variables (Pallier et al., 2002). This analysis essentially focuses on the *confidence* part of overconfidence, and documents correlations within-elicitation, across domain. Blais, Thompson, and Baranski (2005) pursued a similar analytical approach, measuring overestimation using 2AQ questions with confidence ranging from 50% to 100%; they found using CFA that a general model of confidence best fitted data from a variety of tasks, and that a distinct confidence factor emerges (in other words, confidence correlated significantly across tasks, rather than being task specific). For further evidence supporting the possibility of a confidence factor, see Kleitman and Stankov (2001) and Stankov and Crawford (1996).

In other words, these results suggest that someone who displays more overconfidence in one type of task is likely to also display relatively more overconfidence in other instances of similar tasks, with different questions and subject matter (i.e., domains). This indicates that a common factor contributes to their overconfidence across the different instances.

1.4.3 Across elicitation stability

Across elicitation stability is a direct test of our second claim—that at an individual-level, overconfidence will be correlated across the different forms. Cases where researchers included multiple elicitations and compared people's overconfidence across them are rare. In one such case, Larrick et al. (2007) found that overplacement in a given subdomain was a strong predictor of overestimation—with a standardized regression coefficient of .62. They propose that overplacement and overestimation are fundamentally related in that they both involve a subjective sense of competence—and differences in their response to task difficulty are driven by cases where difficulty affects accuracy more than confidence. In another test comparing two elicitations, Muthukrishna et al. (2018) found that although overprecision was correlated with

people's perceptions of their own percentile of performance, such that greater overprecision predicted higher perceived percentiles, it was not correlated with *overplacement* when measured as the difference between people's estimated percentile and their actual percentile. As we have noted earlier, this former measure is the *confidence* component of overplacement, rather than *overplacement*. In other words, these researchers did not find a correlation across overprecision and overplacement elicitations. Notably, a standard deviation based measure of overprecision was used, rather than the more traditional confidence minus hit rate, and the samples ranged across European Canadian, East Asian Canadian, Hong Kong Chinese, and Japanese respondents.

Overall, the Larrick et al. (2007) paper provides some initial evidence for our second claim, but the Muthukrishna et al. (2018) paper's null result suggests a more complete investigation is required to understand the relationships between the different forms of overconfidence.

1.5 Understanding of the nature of overconfidence

Before continuing with our review, we pause to propose the conceptualization of overconfidence we reached in light of the evidence reviewed so far. We argue that there are both theoretical and empirical reasons to believe that different forms of overconfidence share a common, dispositional component that would lead them to be related. On the theoretical side, research into individual differences in reasoning suggests that those who engage in biased search processes in one context may be likely to do so in another (e.g., Stanovich & West, 1998)—and such processes are a key antecedent of overconfidence effects. We also described three different types of data supporting the idea that there may be a common component to overconfidence (split-sample, across domain, across elicitation). Yet there are undoubtedly situational effects that impact overconfidence, and we delineated a series of results that could suggest that the three

forms of overconfidence we introduced are distinct. In the following paragraphs, we attempt to reconcile these seemingly disparate results.

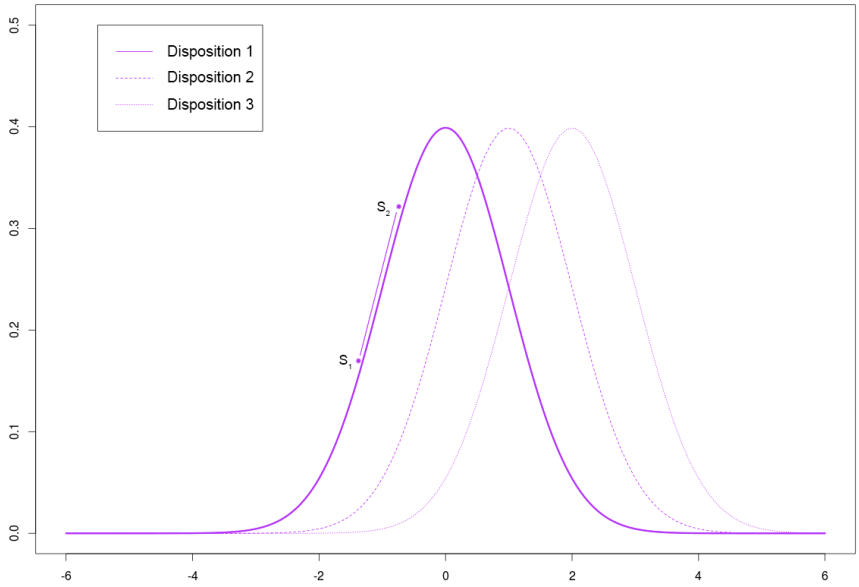

Figure 1. A model of overconfident beliefs.

When looking at an individual, we propose that the degree of overconfidence observed is determined by: i) a dispositional, core component, ii) situational factors, and iii) elicitation effects. Consider Figure 1. People's dispositional overconfidence determines the average level of overconfidence in the distribution—Disposition 2 has a higher average level of overconfidence than Disposition 1; and Disposition 3 has higher *dispositional* overconfidence still. This is driven by the core overconfidence factor that we are primarily concerned with in the present research. Yet there are also strong and vitally important situational effects. These lead to movements *along* the curves. For example, consider S_1 and S_2 in Figure 1. In S_2, situational factors have led to elevated levels of overconfidence relative to S_1. This could be due to a psychological state (such as positive affect) or elements of the task (such as a favored domain, or a change in task difficulty). It could be the case that situational factors overwhelm the difference in dispositions—

in positions along the curve where greater overconfidence is exhibited by Disposition 1 than Disposition 2, these situational factors have proved more critical than the dispositions.

Finally, on top of these two features (the disposition determining the curve, and situational factors determining movements along it), we must also consider the effect of the elicitation. The elicitation is the way we observe the overconfident beliefs. Any one elicitation does two things: i) provides an imperfect measure of the psychological state that is a point drawn from a wider distribution of beliefs (e.g., S_1), and ii) engages other factors that are unrelated with the pure overconfidence observed in Figure 1. To further illustrate the first point, consider that each elicitation has a different mapping between the model of overconfident beliefs and the observed degree of overconfidence. For example, for a given set of beliefs, eliciting 95% confidence intervals is likely to elicit greater overconfidence than eliciting 60% confidence intervals. This is an artefact of the elicitation itself, rather than reflecting anything different about the underlying beliefs. This will affect the levels of overconfidence observed in the measurement, but not necessarily the ordinal ranking of people's overconfidence. Demonstrating the second point, overplacement both captures overconfidence and may engage people's feelings of competition, due to its relative component. The resulting measure is a mixture of the pure measurement of overconfidence and these additional factors.

To put it differently, any time that we observe a measure of overconfidence, we are seeing three components: the trait (dispositional) component T, the situational effect S, and the effect of the form of overconfidence elicited, F, and an error term e:

$$OC = b_1T + b_2S + b_3F + e$$

In the present work we argue that there is a dispositional component (T) in addition to the established effects of situational factors (S)[6] and elicitations (F), but we emphasize that all three of these components are critical in determining the degree of observed overconfidence.[7] This understanding of overconfident beliefs is able to reconcile various findings in the literature, as we describe next. For example, the dispositional and situational components allow for the aggregate effect of factors such as task difficulty while also maintaining the role of dispositional factors that lead to people being more or less overconfident across situations (e.g., a tendency to engage in biased search processes). Our proposal also distinguishes between two parts of each elicitation's measure: a component that accesses people's overconfident beliefs, and a part that is specific to the elicitation which may evoke other aspects of psychology. In cases where an outside behavior (e.g., market entry) is associated with one elicitation (e.g., overplacement) but not another (e.g., overestimation), this could be explained by the elicitation evoking additional psychology (e.g., overplacement may also engage a competitive element). Thus, different elicitations having different relationships with outside variables may reflect the impact of the form (F) component of the above equation, rather than showing the absence of the dispositional, trait (T) component.

We argue that measuring multiple elicitations of overconfidence is important to reliably measure *dispositional* overconfidence—with a single elicitation one could not distinguish between elicitation effects and the dispositional component, which could confound establishing robust relationships between dispositional, core overconfidence and other variables. Further, it could be the case that each elicitation accesses a distinctive part of the dispositional component of

---

[6] These situational factors include stable factors (e.g., the difficulty of the task) which affect everyone, and momentary factors (e.g., variations in individuals' moods)

[7] A full model might further specify person interactions with components S and F. It is possible (and likely) that different individuals respond to different situational and elicitation-based factors differently, and a full model would account for this.

overconfidence. As a result, it is important to ask more than one elicitation to gain a full picture of the nature of *core overconfidence* and to mitigate elicitation effects. To find evidence for our proposed dispositional component to overconfidence, it is also necessary to observe overconfidence over time to mitigate situational effects. We explore both possibilities in the following chapters.

All of this leads to our proposal that at an individual respondent level, overconfidence will be positively correlated across domains and elicitations, resulting from a dispositional component of overconfident beliefs. Situational factors and the artefacts of different elicitations mean that the average levels of observed overconfidence will vary across tasks and time, but we propose that people's *rank orders* within those tasks will be correlated. In line with this logic, our first research question is the following:

RQ1: Are the different forms of overconfidence empirically related? Is there a stable, core component underpinning different measurements?

In asking this question, we hope to investigate whether there is a dispositional component to overconfidence that individuals express across elicitations, which can be measured in a way that minimizes dependence on the idiosyncrasies of particular elicitations but will correlate with different cases of elicitations. The presence of such a *core* overconfidence opens the possibility of finding robust relationships between overconfidence and individual differences in self-views and cognitive style, which may otherwise have been impeded by the roles of elicitations and situational effects. We consider the evidence for such relationships—the fourth type of evidence for individual differences in overconfidence that we introduced (*correlation with traits*)—in the next section.

1.6 Individual difference measures and core overconfidence (correlation with traits)

At least two factors have made it difficult to establish the relationship between overconfidence and individual difference measures, i) the fact that any one measure of overconfidence will be affected by the elicitation, domain, task difficulty, and other situational factors, and ii) the belief that situational effects overwhelm the importance of individual differences *in* overconfidence (Moore & Dev, 2018). The former makes interpreting extant research difficult: if an individual difference measure's correlation with different forms of overconfidence varies significantly across a collection of relatively small samples, one might conclude that the true relationship could be null. The latter has likely stifled people's enthusiasm for such research. Overall, the inconsistency in the correlations between overconfidence and individual difference measures has led some to argue that there is weak evidence for stable individual differences in overconfidence (Moore & Dev, 2018; Moore & Swift, 2011).

We propose that there may be relationships between individual difference measures and what we call *core overconfidence*, which have been obscured by the details of different domains and elicitations. By measuring core overconfidence in a way that depends less on such details, we hope to gain clarity regarding these relationships. It is still possible that there are domain-individual interactions that affect core overconfidence: as someone's knowledge varies across domains this could influence different forms of overconfidence. Yet these may average out, as long as different domains are presented. Before developing our second research question, we start by summarizing what has been found so far regarding the relationship between individual differences and overconfidence.

The most notable finding in the literature is that narcissism is associated with greater overconfidence. This positive association has been shown in different situations; Campbell,

Goodie, and Foster (2004)[8] found that narcissism was associated with greater overestimation, whereas John and Robins (1994) showed narcissism predicted people's bias in their perceptions of their own performance in a managerial group-discussion task, and Ames and Kammrath (2004) found narcissism predicted self-estimated ability at interpersonal perception more effectively than actual competence in such tasks. Further, narcissists *act* on their overconfidence: Those who overestimate their performance bet money on this estimation, resulting in greater losses (Campbell et al., 2004), and overconfidence estimated using a Likert scale measure predicted ignoring expert advice and making impulsive decisions (O'Reilly & Hall, 2021).

It is specifically *grandiose* narcissists, rather than vulnerable ones (Wright, 2016; O'Reilly & Hall, 2021) who are more prone to making overconfident errors. This is the form of narcissism that is characterized by high self-esteem, personal superiority, and entitlement (Miller et al., 2017). This raises the possibility that these traits (e.g., higher self-esteem) might also be associated with overconfidence. In contrast, vulnerable narcissists display a fragile self-concept, typified by anxiety and low self-esteem (Ackerman et al., 2017). Worryingly, Macenczak et al. (2016) find that at high levels of narcissism, there are multiplicative positive effects of grandiose narcissism and power on overestimation. Narcissists who are likely to seek power may be the most corrupted by it, at least in terms of overconfidence. In a further piece of evidence, Paulhus et al. (2003) find that narcissists engage in a more overclaiming—overstating one's familiarity with things—which is related with overconfidence.

Research has also found some evidence for overconfidence's relation with other individual difference measures. Schaefer et al. (2004) report a positive association between

[8] We depart from Moore & Dev's (2018) categorization of the forms of overconfidence—in our typology, overestimation is captured by differences between people's subjective confidence measured for each item and people's performance across a group of items. Overprecision is captured by people providing interval estimates that do or do not include the true population values. In Study 2 we introduce an alternative way of measuring overestimation.

extraversion and overestimation, which is supported by the positive relationship Pallier et al. (2002) found between two extraversion-related constructs (proactiveness and activity) and overestimation. Furthermore, the specific facets of narcissism that Macenczak et al. (2016) find are associated with overestimation are the subfactors associated with extraversion (Leadership/Authority and Grandiose/Exhibitionism).

In addition to self-view variables, others have looked for the relationship between cognitive dimensions and people's displayed overconfidence. Stanovich and West (1998) have found cognitive ability modestly reduces people's susceptibility to overestimation, though Blais et al. (2005) notably failed to relate cognitive style variables to overestimation (including need for cognition, person need for structure, and personal fear of invalidity). Bruine de Bruin et al. (2007) used an overestimation paradigm to create a measure of calibration calculated as, "one minus the absolute difference between mean confidence and percentage correct across items so that higher scores reflect better performance" (p. 942). A higher score on this measure was positively associated with using decision rules, consistency in risk perception, resistance to framing, and recognizing social norms. In other words, individuals who were better calibrated in their confidence judgments also exhibited other normative decision making behaviors. They also showed that calibration was positively associated with education and measures of intelligence. This evidence supports the idea that cognitive style factors will be associated with overconfidence.

There is some additional research considering the relation of confidence or overconfidence with additional individual difference measures. These findings are generally only tested in single studies, with measurements of overconfidence that do not conform to our typology. For further reading, see Olivares (1993), Ramanaiah and Goldberg (1977), Wolfe and Grosch (1990), and Cutler and Wolfe (1989).

Overall, though there is some work relating individual difference measures to overconfidence, the studies elicit overconfidence in different ways, and many associations are weak or inconsistent. Our intention to measure core overconfidence—that which is not associated with a specific elicitation or domain—opens the possibility that there may be stronger associations between this variable and individual difference measures, as it constitutes a more reliable measure of overconfidence. This is an advantage of our approach relative to measuring a single form of elicitation. As a result, our second research question is:

RQ2: What individual difference measures are associated with core overconfidence?

The main takeaway from extant literature is that we should expect to find a positive association between narcissism and overconfidence. There is also some suggestion that extraversion might be positively related with overconfidence, and cognitive ability negatively. The evidence for other relationships is more equivocal. Consequently, we are interested in a wide range of individual difference measures, which can be divided into two categories: those relating to self-views, and those focused on cognitive style. Self-view variables capture how people perceive their own abilities and general performance, whereas cognitive style variables emphasize their approach to thinking about things (including CRT and BNT, which are tests rather than self-reports). We divide some example measures into these two categories below (Table 2), each with a brief explanation of what they measure.

Table 2: Example individual difference measures categorized

Self-view variables	Cognitive style variables
Narcissistic personality (Ames et al., 2006) - Grandiose yet fragile sense of self; need for admiration	Intellectual humility (Leary et al., 2017) - Degree to which recognize beliefs may be wrong
Optimism (Scheier et al., 1994) - Dispositional tendency to hold positive outlook	Cognitive Reflection Test (CRT; Frederick, 2005) - Propensity to engage in deliberative thinking
Self-efficacy (Chen et al., 2001) - Belief in ability to execute actions	Berlin Numeracy Test (BNT; Cokely et al., 2012). - Competency with numbers
Big Five (Gosling, Rentfrow, & Swann, 2003) - Extraversion, agreeableness, conscientiousness, emotional stability, openness to experience	Maximizing vs. satisficing (Diab et al., 2008) - Tendency to attempt to maximize one's outcomes
Dark Triad (Jones & Paulhus, 2014) - Narcissism, psychopathy, Machivellianism	Actively open-minded thinking (Baron, 2019) Disposition to be fair to different conclusions that oppose initially favored ones

Both self-view variables and cognitive style variables could plausibly influence the level of core overconfidence exhibited. Self-view variables could lead to people engaging in motivated search processes that underestimate variability, whereas different cognitive styles may produce more or less exhaustive search procedures, which could also lead to underestimating variability. Moreover, self-view variables may enhance or depress general confidence, whereas cognitive styles could determine people's preferences for calibration. Given that overestimation and overplacement have a clearer subjective confidence component (both rely to some degree on self-reports of performance), we might expect that these forms of overconfidence would be more strongly related with traits capturing people's self-views than overprecision. Conversely, overprecision requires people to adjust their intervals outwards to account for uncertainty in their

knowledge and in the world. The propensity to do so may be more strongly associated with cognitive style variables.

The existing literature suggests that self-view variables (e.g., narcissism) may be more important in determining overconfidence than cognitive style variables, but this is a relatively unexplored question, as few studies have included cognitive style measures. By including a wide range of individual difference measures and extracting core overconfidence that is not dependent on the specifics of its elicitation, we hope to offer a more systematic investigation of its relationship with individual difference measures, helping us to investigate our third research question:

RQ3: Are self-view or cognitive style individual difference measures relatively more important in predicting core overconfidence

Chapter 2: Study 1

2.1 Preface

In our literature review, we introduced three different forms of overconfidence: overestimation, overplacement, and overprecision. We argued that a dispositional component—which we call core overconfidence—could contribute to each of these three forms. Put simply, we argue that people who are more overconfident in one task are likely to be more overconfident in other tasks, even in different time periods. In Study 1, we collected evidence regarding the relationship between overconfidence measures across domains and method of elicitation, and over a long time horizon.

We constructed 12 tasks that measured overconfidence, covering each of the three forms across four domains (geography, history, science, literature). All participants answered these 12 measures, and we tested for positive correlations among them. Positive correlations would suggest that a core overconfidence component may contribute to each case of observed overconfidence. Specifically, we tested the correlations among these measures of overconfidence to test the association of overconfidence i) within-elicitation, across domain (i.e., is overestimation in geography related to overestimation in history), ii) within-domain, across elicitation (i.e., is overestimation in geography related to overplacement in geography), and iii) across domain and elicitation (i.e., is overestimation in geography related to overplacement in history).

We next derived *core overconfidence* by using mixed effects modelling to extract a measure of overconfidence that was less associated with specific domains and elicitations. Core overconfidence essentially captured—relative to the average level of overconfidence in each of the 12 domain-elicitation measurements—what was a participant's average overconfidence?

We validated this measurement by testing whether core overconfidence estimated with a subset of the overconfidence measures could predict a holdout sample of overconfidence measures. Next, we tested whether core overconfidence predicted a new form of self-enhancement: overclaiming familiarity with fabricated items (Paulhus et al., 2003). Finally, we tested whether core overconfidence was associated with a battery of individual difference measures.

The first experimental session was in December 2019. We discuss the results from that session at length in the following section. In order to test the stability of the measured characteristics over time, we launched an additional experimental session in May 2022 where we recruited the same participants to complete the survey again. We report the correlations between each of the measures at time 1 and time 2 to seek evidence regarding whether core overconfidence exhibited reliability across a long time horizon. Substantial test-retest reliability over this window would support our claim that there is a dispositional component to core overconfidence.

2.2 Methods

2.2.1 Participants

We recruited 200 participants from Amazon's Mechanical Turk, who were paid $4 for completing the study. The sample was 41.5% women with a mean age of 37 years. Of the 200 respondents, some did not return complete datasets (typically missing some items from the individual difference measures). As a result, though our main analyses concerning the overconfidence measures contain the full 200 participants, some subsequent analyses omit those who did not respond to the pertinent measures (the number of complete datasets was 187).

2.2.2 Materials

We elicited overconfidence in three different ways—using two-answer-questions (2AQ) to measure overestimation (OE), elicitations of people's estimated quantile in the performance distribution to measure overplacement (i.e., better than average effects, termed OPL), and with confidence intervals to measure overprecision (OPR). Additionally, we tested respondents' propensity to overclaim their knowledge of an area (Paulhus et al., 2003). Paulhus and colleagues introduced overclaiming as a means to measure self-enhancement independently of ability—everyone is equally unfamiliar with made up items. If there is a core overconfidence factor, we might expect that this would be related to such self-enhancing behavior. To design each of the four tasks (overestimation, overplacement, overprecision, overclaiming), we had to specify items.

In order to minimize dependency on any one domain, we selected four domains to use for each type of elicitation. These domains were; geography, history, literature, and science. Within these domains, we followed a procedure to select items. We describe first our rationale, and then describe the procedure.

One critique of overconfidence research has been that particular item pairings distort the likelihood of someone exhibiting overconfidence (Gigerenzer et al., 1991). For example, consider a respondent who is asked which of two countries is more populous. If the respondent was asked to choose between India and Switzerland, they would likely exhibit less overconfidence than if asked between Belgium and the Czech Republic. Given the effect of such choices on the resulting conclusions, it is important to randomly sample items from a domain, rather than cherry picking examples (Klayman et al., 1999).

The OE measure required 20 items, the OPL required 10, the OPR required 10, and the overclaiming measure required 8. We selected the items for our study with the following procedure: i) generate a list of 50 items in the relevant domain, ii) draw the items required for each elicitation without replacement[9], iii) for OE, randomly draw the comparison pairs. This generated the questions for each domain and elicitation in a way that mitigated concerns that the item selection would inflate or distort the observed overconfidence.

[9] The order was overclaiming, overestimation, overplacement, overprecision.

2.2.3 Procedure

Upon entering the study, participants were asked to remember the number 94. They had to enter this number on the next page. Those who failed to do so correctly were removed from the study before providing any data. Respondents then provided informed consent, before being asked to answer based on their knowledge, rather than by using the internet to research their answers.

The first four blocks contained the overestimation elicitation of overconfidence (Klayman et al., 1999). For each domain, participants had to answer 10 questions and indicate their chance of being correct in each question using a radio button scale from 50 to 100% in increments of 5%. For the geography domain, participants were asked which of two countries had a larger population; for history, participants were asked which of two historical events occurred first; for literature they were asked which of two novels was published first; and for science they were asked which of two discoveries was made first.

The next four blocks contained the overplacement elicitation of overconfidence (Larrick et al., 2007). Participants were asked 10 questions and then asked to estimate their percentile in the performance distribution. Specifically, respondents were told, "*Think about the questions you just answered. Now think about how many you got right. Compared to other participants who are taking this survey, what percentile do you think you fall in, in terms of how many you got right? If you think you are better than 90% of the participants,*

you would answer 90%. If you think you are better than 10% of the participants, you would answer 10%." For the geography domain, participants were asked to estimate a population; for the other three domains they were asked to estimate the date of an event, book, or scientific discovery. After answering the BTA items, participants responded to the four overclaiming blocks (Paulhus et al., 2003). The overclaiming task asked participants to indicate their familiarity with 10 objects on a Likert scale; of which 8 were real and 2 were fabricated.

In the final elicitation, participants provided confidence intervals in a measurement of overprecision. For a quantity—population for geography; dates for the other domains—participants were asked to provide a lower and upper estimate. For example, we asked participants for, "*a lower estimate such that you are 90% sure that Norway does not have fewer people than that*", and "*an upper estimate such that you are 90% sure that Norway does not have more people than that*".

In sum, we asked participants to respond to three different elicitations of overconfidence (OE, OPL, OPR) and an overclaiming task. We did not randomize the order of these blocks to maximize ease for participants; but did randomize the order of items within blocks (except for the OE items because answers and confidence were elicited separately for each item).

After the overconfidence measures, participants provided responses to a series of individual difference measures. These were; intellectual humility (Leary et al., 2017),

need for closure (decisiveness facet), need for closure (closed mindedness facet; Kruglanski et al., 1993), the narcissistic personality inventory (Ames et al., 2006), optimism (Scheier et al., 1994), self-efficacy (Chen et al., 2001), desirability of control (Burger & Cooper, 1979), and the dominance facet of extraversion (Wiggins et al., 1988). The order of these measures, and the items within the measures, was kept constant. Finally, participants provided demographic information (age, gender, education, race).

2.2.4 Coding the data

Below, we briefly describe how we coded people's raw responses into our measures of overconfidence. For the OE items, each item had a single correct answer (i.e., one of the choices was correct). We calculated how many items each respondent got right in a domain, and their average confidence across their responses. We then subtracted the percentage of correct responses from the average confidence to compute overestimation.

For the OPL items, each item had a specific correct answer (e.g., James Chadwick discovered neutrons in 1932), but we used mean absolute deviation (MAD) to score the accuracy of participants' responses, rather than a count of correct answers. Within each domain, we averaged the absolute deviation of each of the respondent's 10 answers from the true value to get an overall measure of their accuracy (MAD). We then ranked the participants within each domain based on this measure to produce an estimate of their quantile in the performance distribution. We then compared this quantile to the

participant's estimate of their performance quantile to calculate a measure of overplacement.

For overprecision we counted how many of a respondent's confidence intervals contained the true value and compared this to the number that should have contained the true value—which was 8, as we elicited 80% confidence intervals. As a result, participants' overprecision was bounded by -20% and 80% (the cases where all 10 and 0 of the intervals contained the true value, respectively).

Though Paulhus and colleagues (2003) analyzed overclaiming using signal detection methods, we opted for a simpler measure. It involves two primary components: the average familiarity with the fabricated items (F) and the average familiarity with the real items (R). The formula we use to compute the measure is F(7 – (R – F)). A higher score on this variable can indicate two things: i) a smaller disparity in familiarity between real and fake items, and ii) a higher general tendency to claim familiarity with fake items. These are both self-enhancement behaviors. We intend to use this as an outcome variable to help to validate our measurement of overconfidence.

Finally, we coded the scales as per their respective coding schemes. The alphas were as follows; intellectual humility (0.84), need for closure decisiveness (0.84), need for closure closed mindedness (0.76), narcissistic personality inventory (0.87), optimism (0.91), self-efficacy (0.93), desire for control (0.80), and the dominance facet of extraversion (0.90).

2.2.5 Results

The core of our dataset is 12 measures of participants' overconfidence—using three elicitation methods across four different domains. In addition to these measures of overconfidence, we also acquired four measures of overclaiming, participants' scores on eight individual difference measures, and demographic information. We begin our analyses by studying the exhibited overconfidence and the relationship between the different measures of overconfidence.

2.2.5.1 Nature of overconfidence

Respondents were on average overconfident (M_{OC} = 10.4%) across all measures, though this average conceals significant variation across elicitation method. Whereas there was some overestimation (M_{OE} = 5.6%) and significant overprecision (M_{OPR} = 35.4%), participants were on average underconfident for the overplacement items (M_{OPL} = -9.8%). Given that the study tasks were quite difficult, this is consistent with "better than average" effects, where experienced difficulty can lead to underconfidence when using people's perceived performance to measure confidence (Burson et al., 2006; Larrick et al., 2007). Moreover, the greater overconfidence observed in overprecision relative to overestimation is consistent with other research (e.g., Klayman et al., 1999).

In sum, the different elicitations yielded different average levels of overconfidence. Yet the average levels of overconfidence across tasks does not tell us anything about the ordering of people within those tasks. Could it be that those who

indicated the highest levels of overplacement also displayed the highest levels of overprecision? We next ask this question by testing for correlations between the different measures of overconfidence. Recall that all participants answered 12 measures of overconfidence: responding to four different domains in each of the three elicitations. We provide the correlations among these 12 measures below in Table 3.

Table 3: The correlations between the 12 measures of overconfidence.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. oeg	9	19											
2. oeh	-0	17	.40** [.28, .51]										
3. oel	5	21	.42** [.30, .53]	.55** [.44, .64]									
4. oes	9	20	.38** [.26, .49]	.65** [.56, .72]	.52** [.41, .61]								
5. oplg	-1	31	.23** [.10, .36]	.30** [.17, .42]	.20** [.07, .33]	.33** [.20, .45]							
6. oplh	-5	34	.27** [.14, .40]	.39** [.26, .50]	.34** [.22, .46]	.32** [.19, .44]	.29** [.15, .41]						
7. opll	-17	37	.14* [.01, .28]	.32** [.19, .44]	.23** [.10, .36]	.34** [.22, .46]	.32** [.18, .44]	.49** [.37, .59]					
8. opls	-16	38	.41** [.28, .52]	.45** [.34, .56]	.37** [.25, .49]	.40** [.28, .51]	.40** [.28, .51]	.56** [.46, .65]	.49** [.38, .59]				
9. oprg	42	30	.26** [.12, .38]	.34** [.21, .46]	.31** [.18, .43]	.28** [.15, .41]	.25** [.11, .37]	.17* [.03, .30]	.19** [.06, .32]	.22** [.08, .34]			
10. oprh	33	29	.39** [.27, .50]	.57** [.47, .66]	.42** [.30, .53]	.46** [.34, .56]	.23** [.10, .36]	.23** [.09, .36]	.25** [.12, .38]	.35** [.22, .47]	.51** [.40, .60]		
11. oprl	29	28	.39** [.26, .50]	.49** [.37, .59]	.39** [.27, .50]	.45** [.34, .56]	.22** [.09, .35]	.21** [.07, .34]	.24** [.10, .36]	.33** [.20, .45]	.50** [.39, .60]	.74** [.68, .80]	
12. oprs	37	25	.33** [.20, .45]	.47** [.35, .57]	.38** [.25, .49]	.38** [.26, .50]	.18* [.04, .31]	.17* [.04, .30]	.17* [.03, .30]	.28** [.15, .41]	.54** [.44, .63]	.71** [.63, .77]	.78** [.72, .83]

Note. The 12 variable names each have an identifier that conveys the elicitation: oe for overestimation, opl for overplacement, and opr for overprecsion. Appended to this is a single letter identifier that conveys the domain; g for geography, h for history, l for literature, and s for science.
Note. ** $p < .01$, * $p < .05$

The correlations between *every* domain-elicitation pair attained statistical significance. Notably, the average correlations within-elicitations among domains were

very high; 0.49 for overestimation, 0.43 for overplacement, and 0.63 for overprecision. Within-domains, the average correlations between elicitation methods were relatively high for all domains, including; geography 0.25, history 0.40, literature 0.29, and science 0.35. The correlations among measures from different domains *and* elicitations can be seen in Table 3.

The two motivating questions of the present research were whether there is a dispositional component to overconfident beliefs and whether the different forms of overconfidence would be related at an individual-level. The correlations in Table 3 provide strong evidence relating to the second question. The evidence so far suggests that the overconfidence is highly correlated across domains and elicitations. This could also indicate a dispositional component, though stability over time is necessary to determine that these relationships are not driven by common situational factors (e.g., positive affect). Regardless, the associations among the different forms of overconfidence raise the possibility of measuring overconfidence in a way that does not depend on particular domains and elicitations to focus on a common component. The following analyses test this possibility.

2.2.5.2 Measuring core overconfidence

We have proposed that any one measure of overconfidence is influenced both by situational effects (e.g., the difficulty of the domain, the respondent's mood on the day), task elicitation effects, and a common factor. Distinguishing between these influences

requires respondents answering more than one overconfidence-based task. One option to mitigate task effects (both from elicitations and situational factors) would be to average several measures of overconfidence ("the simple average"). We pursued a slightly different approach, analyzing participants' overconfidence using mixed effects models. To do so, we first pivoted each participant's data to long form, where each row of data was a separate measurement of overconfidence. We then computed the mean and standard deviation for each domain-elicitation pairing, and used this information to standardize each overconfidence measure *within* domain-elicitation (subtracting the mean overconfidence and dividing by the standard deviation). Using this data, we estimated a model where the observed overconfidence was the dependent variable and the independent variables included a grand intercept and random intercepts nested within individuals. These random intercepts were the measure of core overconfidence. The resulting measure is very similar to the simple average: it essentially takes the simple average and shrinks it towards the grand intercept (which was 0 due to standardization) to correct for unreliability in the measures. As a result, it should be a better predictor of out of sample measures.

Due to our approach, we were able to test how reliable our core overconfidence measurement was. We analyzed this using the pooling factor for the multilevel model, which indicates the extent to which the individual-level random intercepts were shrunk towards the grand intercept (Gelman & Hill, 2006). We will report 1 – w, where w is the

pooling factor: based on this measure, a score of 0 indicates complete pooling (i.e., the individual-level random intercepts are not helpful in estimation) and a score of 1 indicates no pooling (i.e., the random intercepts are not shrunk at all to the grand intercept). In our data the grand intercept is 0, as we standardized our measures within domain-elicitation, which means that 1 – w can be computed as the ratio between each participant's random intercept and the simple average of the twelve overconfidence measures for that participant. Across all respondents, this ratio was 0.876. This indicates little pooling and high reliability of our measure of core overconfidence.

2.2.5.3 Association between core overconfidence and specific contexts

Core overconfidence was strongly correlated with each of the 12 separate measures of overconfidence, but this is logically necessary given these measures are its constituent parts. To further validate this measure, we dropped one elicitation at a time from the estimation and used it as a holdout sample. Essentially, we estimated core overconfidence using 2/3 elicitations (e.g., overplacement and overprecision) and used the resulting core overconfidence estimate to predict each of the four domain-level measures in the held out elicitation (e.g., overestimation).

When estimating core overconfidence without including the overestimation measures, the resulting core overconfidence correlated significantly with each of the overestimation domains; geography ($r = 0.46$, $p < .001$), history ($r = 0.63$, $p < .001$), literature ($r = 0.50$, $p < .001$), and science ($r = 0.56$, $p < .001$). The corresponding

correlations when dropping overplacement ($r = 0.34$, $r = 0.36$, $r = 0.32$, $r = 0.48$, all $p < .001$) and overprecision ($r = 0.37$, $r = 0.54$, $r = 0.51$, $r = 0.44$, all $p < .001$) were also positive and significant. In short, core overconfidence predicted measurements of overconfidence that were not included in its construction.

This suggests that there is a common component that affects the observed overconfidence across the different elicitations. This could be a dispositional component, or it could reflect the fact that situational factors affect all three elicitations simultaneously. We later test the stability of core overconfidence over time to distinguish between these competing explanations. Regardless, we find robust associations of overconfidence across different elicitations, supporting our second primary claim.

2.2.5.4 Core overconfidence and overclaiming

The analyses so far show there is a common core overconfidence that underlies different measures of overconfidence, but its predictive power has not yet been tested in a new context. It could be the case that the observed consistency across the forms of overconfidence is attributable to an elicitation effect in their response scales—each time eliciting numeric estimates and/or confidence estimates.[10] To address this concern, we

[10] We do not find this argument especially compelling given the large differences between the overestimation/overplacement and overprecision elicitations, but the association between core overconfidence and a new measurement with a new response format provides more evidence that the associations across forms of overconfidence are not *just* elicitation effects.

used our measures of overclaiming to ask the question: Can core overconfidence predict people's self-enhancement in a new task with a different response format?

In the overclaiming paradigm, people indicated their familiarity with real and fabricated objects and events. For example, respondents were asked how familiar they were with the Louisiana Purchase and Frank Smith's election as U.S. President. People should indicate greater familiarity with the Louisiana Purchase (a real event) than the election of a fake president, but those who are generally prone to inflated beliefs regarding their own knowledge might indicate similar levels of familiarity. Consequently, both generally indicating greater familiarity with fake items and a smaller gap in people's average familiarity between real and fabricated objects are forms of self-enhancement, which we might expect to be related to core overconfidence.

We tested the correlations between core overconfidence (now, and in all subsequent analyses, estimated with the *full* 12 measures) and overclaiming in each of four domains (geography, history, literature, science). All of these correlations were significant and positive ($r = 0.70$, $r = 0.72$, $r = 0.55$, $r = 0.59$, all $p < .001$). In other words, those who scored higher in core overconfidence were more likely to overclaim, displaying a different form of self-enhancement.

In sum, both by using hold out samples from the overconfidence measures, and a new, related behavior (overclaiming), we showed that core overconfidence is associated with a number of different instances of self-enhancement. Our next question is whether

this core overconfidence is also associated with established individual difference measures.

2.2.5.5 Overconfidence and individual differences

Of our individual difference measures, core overconfidence was significantly correlated with intellectual humility (r = -0.35, p < .001), the decisiveness facet of need for closure (r = -0.18, p = .012), the closed mindedness facet of need for closure (r = 0.15, p = .032), narcissism (r = 0.58, p < .001), and the dominance facet of extraversion (r = 0.41, p < .001). These correlations were all in the expected directions, with the exception of decisiveness, which it appears was negatively associated with overconfidence. This could be explained by people with higher scores on this facet for need for closure feeling an urge to make a decisions swiftly to mitigate their own decision uncertainty, but still being aware of the uncertainty associated with making such swift decisions. We will have to replicate this result to have confidence in it. There were very strong correlations between core overconfidence and intellectual humility, narcissism, and the dominance facet of extraversion. Notably, core overconfidence did not significantly correlate with optimism (r = -0.01, p = .887), self-efficacy (r = 0.09, p = .219), or desire for control (r = -0.10, p = .146).

Overall, the data suggest that core overconfidence is associated with other stable individual differences—both in terms of self-views (e.g., narcissism) and people's cognitive styles (e.g., intellectual humility). Though the stronger correlations

were with the self-view variables (e.g., narcissism), these sets were by no means exhaustive and so further work is required to determine which sorts of traits are more strongly associated with core overconfidence. We expand on this further in a supplemental study in Appendix B.

2.3 Second experimental session

In May 2022, we invited all 200 of the participants of Study 1 to complete the survey again, two and a half years after first completing it. We did not explicitly communicate that it was the same survey, but rather offered $7 to complete a 40 minute task to this group of participants. We were able to capture a second wave of data for 85 participants. In Table 4, we show the correlations between each of the traits we measured across this long time window for these 85 participants.

Table 4: Correlations among Time 1 (rows) and Time 2 (columns) measures of each variable.

	Time 2 measures								
Variable	Core OC	IH	Decisive	Closed	NPI	Optimism	Efficacy	Control	Dominance
Core OC	.77** [.67, .85]	-.36** [-.53, -.16]	.05 [-.17, .26]	.03 [-.18, .25]	.38** [.18, .55]	.10 [-.11, .31]	.11 [-.11, .31]	-.05 [-.26, .17]	.18 [-.03, .38]
IH	-.32** [-.50, -.12]	.73** [.62, .82]	-.13 [-.33, .09]	-.44** [-.59, -.25]	-.26* [-.45, -.05]	-.10 [-.30, .12]	-.02 [-.23, .20]	.17 [-.04, .37]	-.22* [-.41, -.01]
Decisive	-.09 [-.30, .12]	.03 [-.18, .24]	.83** [.75, .89]	-.04 [-.25, .17]	.10 [-.11, .31]	.50** [.32, .64]	.51** [.33, .65]	.42** [.22, .58]	.40** [.21, .57]
Closed	.04 [-.18, .25]	-.39** [-.55, -.19]	.17 [-.05, .37]	.79** [.69, .86]	.16 [-.06, .36]	-.11 [-.31, .11]	-.13 [-.34, .08]	-.15 [-.35, .07]	-.00 [-.21, .21]
NPI	.50** [.31, .65]	-.33** [-.51, -.12]	.09 [-.14, .30]	.10 [-.12, .32]	.83** [.75, .89]	.06 [-.16, .28]	.16 [-.06, .37]	.17 [-.06, .37]	.40** [.19, .57]
Optimism	.01 [-.21, .22]	.04 [-.18, .25]	.47** [.29, .62]	-.22* [-.41, -.01]	.09 [-.13, .29]	.89** [.84, .93]	.64** [.49, .75]	.27* [.06, .45]	.53** [.36, .67]
Efficacy	.01 [-.20, .23]	.12 [-.09, .33]	.51** [.34, .65]	-.25* [-.44, -.04]	.11 [-.11, .31]	.71** [.59, .81]	.85** [.77, .90]	.33** [.13, .51]	.57** [.40, .69]
Control	-.16 [-.36, .05]	.13 [-.08, .34]	.35** [.15, .53]	-.20 [-.39, .02]	.22* [.00, .41]	.23* [.02, .42]	.32** [.12, .50]	.77** [.66, .84]	.39** [.20, .56]
Dominance	.22* [.01, .41]	-.06 [-.27, .15]	.45** [.27, .61]	-.20 [-.39, .02]	.45** [.26, .61]	.55** [.38, .68]	.58** [.42, .71]	.53** [.36, .67]	.85** [.77, .90]

Note. The variable names in Table 4 are as follows; Core OC (core overconfidence), IH (intellectual humility), decisive (the decisiveness facet of need for closure), closed (the closed-mindedness facet of need for closure), NPI (narcissistic personality inventory), optimism, efficacy (self-efficacy), control (desire for control), and dominance (the dominance facet of extraversion). These are the measures described earlier in the results session.

Note. The way to interpret the rows and columns is the following: consider the second row in the first column (a correlation of -.32). This is the correlation between intellectual humility at time 1 and core overconfidence at time 2. The second column in the first row (a correlation of -.36) is the opposite—the correlation between core overconfidence at time 1 and intellectual humility at time 2.

Note. * $p < .05$; ** $p < .01$

Notably, core overconfidence displayed a strong correlation over time ($r = 0.77$, $p < .001$). The magnitude of this correlation was similar to the other measures, which are established measures of stable, individual differences. These data provide strong

evidence that core overconfidence is stable even across a very long time period, supporting the idea that there is a dispositional component to overconfident beliefs.

In Table 5, we break out the test-retest stability of overconfidence by form. Overestimation ($r = 0.64$, $p < .001$), overplacement ($r = 0.58$, $p < .001$), and overprecision ($r = 0.76$, $p < .001$) all exhibited significant reliability over a long time horizon. Core overconfidence at time 1 also significantly predicted all of the forms of overconfidence at time 2. The correlations between overplacement at time 2 and the time 1 measures of overconfidence were somewhat lower, which could reflect reduced reliability in the second sample due to it being smaller.

Table 5: Correlations among Time 1 and Time 2 measures of the forms of overconfidence.

	Time 2 measures			
Variable	Core OC	OE	OPL	OPR
Core OC	.77** [.67, .85]	.61** [.45, .72]	.29** [.08, .47]	.66** [.52, .77]
OE	.59** [.43, .71]	.64** [.49, .75]	.16 [-.06, .36]	.43** [.24, .59]
OPL	.58** [.42, .71]	.34** [.13, .51]	.58** [.42, .71]	.26* [.04, .44]
OPR	.59** [.43, .72]	.43** [.24, .59]	-.03 [-.24, .19]	.76** [.66, .84]

Note. We adopt the same nomenclature as before—OE is overestimation, OPL overplacement, OPR overprecision, and OE (alternate) is the new measure of overestimation.

Note. The way to read the rows and columns is the same as Table 4: the correlation between time 1 OE and time 2 core OC is $r = .59$; and the correlation between time 1 core OC and time 2 OE is $r = .61$.
Note. * $p < .05$; ** $p < .01$

Given the relatively high attrition rate, it is important to check how representative the 85 participants who returned are of the wider sample (and indeed, how likely they were to be the same people, given it is possible that the MTurk accounts could have changed hands). To check these two points, we did three things: i) descriptively compared the levels of each variable across the whole sample and the subsample, ii) cross-tabulated people's genders across their two reports of demographic information, and iii) tested the correlations among their reported ages at time 1 and time 2.

The returning sample was older (M_{age} = 41.8y vs. M_{age} = 37.1y, $p = .006$), which can partially be explained by the 2.5 year time delay. It was also somewhat less educated (M_{educ} = 2.86 vs. M_{educ} = 3.10, $p = .021$), and lower in narcissism (0.161 vs. 0.236, $p = .005$) and dominance (4.32 vs. 4.75, $p = .025$). The subset did not differ from the complete sample for the other traits (e.g., intellectual humility). The gender cross-tabulations showed only 3 off-diagonal participants, and the correlation among reported ages was $r = .98$ ($p < .001$).

In sum, the two-stage sample appeared to be both representative of the full sample of Study 1, and high quality in terms of its consistency. It seems unlikely that the

strong test-retest reliability of core overconfidence could be driven by selective attrition—if anything, the slightly lower narcissism and dominance scores of the returning sample might be expected to attenuate the stability of overconfidence over time. Yet to further verify that this effect was not driven by attrition, we include Study 2 which had a much lower attrition rate between its two waves.

2.4 Discussion

Study 1 provided several key results. Notably, we observed strong positive correlations between measures of overconfidence across elicitation method and domain, suggesting there may be a common component that influences overconfidence across measures. Using mixed effects modelling to extract the overconfidence that was not associated with a particular elicitation method or domain, we found evidence for a "core overconfidence" factor that was strongly associated with holdout samples of overconfidence measures, people's propensity to overclaim familiarity with items, and individual difference measures, particularly narcissism, dominance, and intellectual humility.

Our measure of core overconfidence exhibited strong reliability over time—with a correlation over a 2.5 year time period commensurate with scales that measure stable personality traits (e.g., desire for control, the decisiveness facet of need for closure). This supports the idea that the common factor underpinning the different measurements of overconfidence has a dispositional component. In short, both of our primary claims were

supported: there appears to be a dispositional component to overconfident beliefs, and the three forms of overconfidence are significantly positively related at an individual level.

There are some limitations of the present work. Though we randomly sampled our items when constructing the overconfidence tasks, a choice made to avoid non-representative cue validities (Gigerenzer et al., 1991; Klayman et al., 1999), these were not sampled from comprehensive lists. This could bias the observed levels of overconfidence, but should not affect the association between different forms of overconfidence. Second, for our overplacement task we instructed respondents to estimate their performance quantile based on the *number* of questions they answered correctly, though we later determined to use mean absolute deviation to evaluate their accuracy. It is possible (though unlikely) that this could have influenced their confidence estimates. We address this latter point in Study 2.

Further, though our measure of the test-retest reliability of core overconfidence had the advantage of being over a *long* time horizon, thus mitigating the role of contemporaneous situational effects (e.g., mood), the end sample across both stages was relatively small. It would be advantageous to also see the stability of core overconfidence over time with a larger sample with a lower attrition rate. We investigate this possibility next—asking participants to complete two measures of core overconfidence with a 1-week time gap between them.

There are also key points remaining that require addressing in future empirical work. For instance, is core overconfidence a product of confidence, accuracy, calibration, or some mixture of them? In the overestimation and overplacement paradigms, there is a clear distinction between respondents' stated confidence levels and the accuracy of the answers they deliver, but in overprecision it is more difficult. In overprecision questions, people indicate multiple estimates that determine an interval. The location of this interval (i.e., does it surround the true value) determines accuracy and overconfidence. Accuracy and overconfidence are collinear in this case, as the appropriate confidence level (e.g., 80%) is fixed by the question. Yet the *width* of the interval also provides information about people's confidence: How wide was the range of values they considered? Moving forward, we use the mean expected absolute deviation (MEAD) of people's intervals as a measure of interval width (adjusting for asymmetry) to measure confidence separately from overconfidence in the overestimation paradigm (see Soll & Klayman, 2004).[11] This facilitates teasing apart the factors that contribute to overconfidence.

[11] Relatedly, note that the overconfidence exhibited in the overprecision estimates is bounded by -20% to 80%, which reiterates the importance of standardizing overconfidence within elicitation methods.

Chapter 3: Study 2

3.1 Preface

In Study 1 we asked people 12 different measures of overconfidence each, and showed that overconfidence was significantly correlated across elicitations and domains. We used these 12 measures to estimate *core overconfidence* that was less associated with the specific domain and elicitations that a respondent answered. The resulting measurements of core overconfidence correlated significantly with holdout samples of overconfidence measures, and responses to a different self-enhancement task (overclaiming). In sum, it appears that different measures of overconfidence share a common factor, and this factor is associated with other forms of self-enhancement. However, when only considering one time period, this factor could be a situational effect. For example, perhaps people who experience positive affect at the time of completing the experiment indicate more positive evaluations across the board. To address this alternate explanation, it is necessary to measure people's core overconfidence across time. We tested the stability of core overconfidence over a long time horizon by recalling the participants of Study 1 to complete a second session, but this was a relatively small sample. In Study 2, we measured core overconfidence in two time periods for a larger group of people with a lower attrition rate across sessions.

Study 2 had two key aims: i) establishing whether the positive correlations among different forms of overconfidence were robust in a new sample, and ii) testing

the reliability of core overconfidence across a one-week time period. Establishing positive correlations in each case would further support the claim that there is a dispositional component to core overconfidence, and consequently that the three forms are empirically related.

In addition to its primary objectives, Study 2 tested several other hypotheses. First, we tested the effects of calculating overestimation in a different way. Moore and Healy (2008) argue that the classical implementation of overestimation confounds overestimation and overprecision, as the confidence one states in a 2AQ judgment reflects one's subjective probability distributions over different quantities. The method that they propose compares the number of questions a respondent reports that they got right to the true number. We included this measure in Study 2 to test whether overestimation is related with other forms of overconfidence when using this alternative implementation.

Second, given that we measured twelve different estimates of overconfidence at two time periods, it is possible to test the stability of different elicitations over time. For example, is overestimation at time 1 more strongly correlated with overestimation at time 2 than the corresponding measures of overplacement are? Third, we included centrality measures for our overprecision measures, allowing us to tease apart the contributions of confidence, accuracy, and calibration to the observed pattern of overconfidence results. Finally, we included nine items we are calling *overconfidence*

innovation (OCI) items with the goal of piloting items that can assess dispositional overconfidence briefly and efficiently. Specifically, we use the OCI items to try to construct a three-item test that meaningfully correlates with core overconfidence.

3.2 Methods

3.2.1 Participants

Four-hundred ninety seven MTurk participants completed the first stage of the study in exchange for $3. This initial sample had an average age of 40.7y and was 54.1% men. All participants provided complete datasets. Exactly one-week after this first data collection ended, we invited all of the participants to return to complete a second experimental session for $5 (see registration: https://osf.io/kzw3x). The study was made available for 5 full days. In total, 442 participants returned (M_{age} = 41.3y, 53.6% men).

3.2.2 Materials

Respondents completed 12 overconfidence measures in each session of Study 2. The overconfidence measures of Study 2 were identical to Study 1, other than three small differences. First, the overprecision items now requested centrality estimates: in addition to asking for a lower estimate and upper estimate, we asked for a median estimate. This median estimate was required to estimate the MEAD measure of

confidence in the overprecision paradigm when allowing for asymmetry.[12] Second, we added a new question to the end of each overestimation block, asking respondents, "*How many of the 10 [domain] questions do you think you got right?*". This was to measure overestimation with a different approach as per Moore and Healy (2008). Finally, we modified the wording of our elicitation of the confidence estimate in the overplacement paradigm to use language that reflected our ranking of people's accuracy by MAD rather than number of questions answered correctly.

In addition to completing these 12 overconfidence measures, respondents completed the same nine *overconfidence innovation* (OCI) items in each session, which were short items aimed at measuring core overconfidence efficiently. The construct these items measure is the same—core overconfidence—but these items aim to capture it in a domain-agnostic manner, with fewer questions. We also included an item (OCI9 below) to test whether respondents had metacognitive awareness of their own overconfidence. This uses the same logic as the "Single Item Narcissism Scale" (Konrath et al., 2014), which found that people's narcissism could be effectively measured by asking them how narcissistic they are. These items are included in full below:

OCI1. One-hundred people are guessing the number of jellybeans in a jar. The closest 10 guesses win $100. How much would you pay to play?

[12] We could have estimated the MEAD using just the lower and upper estimates, but it would have required assuming symmetry and normality. We included the median estimate so we could use beta distributions to estimate respondents' subjective probability distribution functions, which allow for skew.

What about if no-one gets to see the jar?
OCI2. You're the CEO of a small confectionary company entering the market for snack bars. What percent of your time would you spend... (0-100%) - Working on your own product - Preparing your marketing campaign - Researching the competitors - Other
OCI3. You're in a large group of people playing a new game, Zoob. You don't know how your ability compares to other people's. You play one game against a randomly selected person and win. What percentage of people do you think you're better than? (0-100%)
OCI4. You're playing a card game with four other people where you ask people questions to learn about their cards. You know that you have 4 clubs in your hand. How likely are you to figure out if you have the most clubs? (from 0-100%)
OCI5. Imagine you're on a committee deciding between 5 different project options. There are 10 members of the committee. You propose Option A. The next three committee members propose Option B. The six other committee members will propose their choices next. When the meeting is finished, what percentage of the choices do you think will be Option A? (from 0-100%)
OCI6. If there's 100 randomly selected people in a room, what's your lower bound for… How many went to college (0-100)
OCI7. If there's 100 randomly selected people in a room, what's your lower bound for… How many own a yellow car (0-100)
OCI8. In vision tests, people are shown flashing dots in different areas of a screen and asked

to press a button indicating that you see them. If you did this test sped up by ten times, of 100 dots, what's the minimum number you think you'd spot? (0-100)
OCI9. How overconfident are you in how you make decisions? (0-10)* *This item we see as distinct from items 1-8 and is included to test people's metacognitive awareness of their calibration. We used a response scale from 0 to 10 to reflect the reduced precision of responses to such a question.

Finally, we also included the narcissistic personality inventory (Ames et al., 2006) in the second experimental session.

3.2.3 Procedure

In the first experimental session, respondents were asked whether they would be willing to complete two surveys—one now, and one in a week's time. If they indicated that they were not, they were removed from the study. Participants also had to successfully answer a comprehension check where they saw the text "472 Printer 124" and were asked to return only the words not the numbers into a text entry box. Failing to return "printer" or "Printer" also resulted in expulsion from the study prior to providing any data.

In the first session, participants then responded to the 12 measures of overconfidence (and the alternate measure of overestimation). These were randomized as follows: i) participants faced the overestimation, overplacement, and overprecision sets of tasks in a randomly selected order, ii) within that order, the domains in each

elicitation were randomly presented. We kept the different domains of the same elicitation in blocks so respondents could stay attuned to the nature of the specific task they are completing. Within the domains, the order of the items was randomized for overplacement and overprecision but not overestimation (as in Study 1; because each response to the overestimation questions required an answer and a confidence judgment). After completing the 12 measures of overconfidence, respondents answered the first eight of the overconfidence innovation items in a randomly selected order, before finally answering the ninth overconfidence innovation item, and providing demographic information.

One-week after the first experimental session was completed, the respondents were invited back to complete a second experimental session. This was made available for five days. In the second experimental session, and respondents were offered $5 to complete the same survey again, with the addition of the narcissistic personality inventory (Ames et al., 2006). The narcissism measure came after the overconfidence measures and before the OCI items. Respondents were not informed that they were completing the same survey again, and the randomized orders of the overconfidence tasks were not held constant across the waves.

3.2.4 Coding the data

The overconfidence measures were coded the same as in Study 1. We also analyzed participants' confidence and accuracy. For overestimation, these were simply

the average confidence they indicated for each of their answers and the percentage answered correctly; for overplacement it was their estimated percentile in the performance distribution and actual percentile (based on mean absolute deviation: MAD); for overprecision, it was slightly more complicated. In our overprecision questions, we estimated the confidence espoused by the respondents' answers by looking at the mean expected absolute deviation (MEAD) associated with their provided interval estimates (Soll & Klayman, 2004). To do so, we elicited three quantiles from respondents: the 10th, 50th, and 90th percentile of their subjective probability density function for an answer. We used these percentiles to estimate the parameters of a beta distribution using the R package 'SHELF' (Oakley, 2021).[13] Using these parameters, we used simulations to compute the expected absolute deviation of the observed outcome from the respondent's median estimate. We averaged these to get the MEAD across the 10 questions in each domain. We computed the MAD of the answers by computing the average absolute deviation between the respondents' median estimates and the true values across the domain. Our measure of respondents' confidence for overprecision was simply the MEAD, and our measure of accuracy the MAD (Soll & Klayman, 2004).[14]

---

[13] The results were consistent if we chose to just use the 10th and 90th percentiles to estimate a symmetric normal distribution's parameters.

[14] There are several ways in which respondents' estimates could deviate from the expected pattern (three distinct, monotonically increasing estimates). First, if participants indicated monotonic but decreasing estimates, these were reversed before being analyzed. Second, if participants indicated estimates that were *all* equal, we could not estimate a beta distribution, but instead entered an EAD of 0. If people indicated other erroneous estimates (e.g., 10th percentile = 50th percentile but not 90th percentile, or other non-

In addition to our primary measures of overconfidence, we also measured overestimation using a different approach. Here we simply compared the number of questions a respondent answered correctly in a domain (e.g., 6) to their estimate of how many they answered correctly (e.g., 8). This would constitute an overestimation of 20%.

3.2.5 Results

Our results can be divided into two main categories: those analyzing the nature of overconfidence, and those testing its stability over time. We start by discussing the results of the first stage of the study, before reporting the stability of different measures of overconfidence over time, and an additional validation that stems from analyzing both stages jointly.

3.2.5.1 Nature of overconfidence

We begin our analyses by observing the levels of overconfidence in each of the elicitations (Table 6). We observed, on average, a small degree of overestimation, a larger degree of underplacement, and a very strong pattern of overprecision. This pattern mirrors the results of Study 1 and broader findings in the literature.

As we noted, we also measured overestimation using a different approach. We include this measure under the label OE (alternate) below. Using this calculation, we

---

monotonicity) these were removed from the observations. If a participant had more than 8 valid estimates, MEAD and MAD were still computed; if they did not, their response for overprecision's confidence measure was made NA.

observed underestimation. This is in line with results in the literature, which finds that people exhibit less overestimation when taking an outside view versus an inside view (Kahneman & Tversky, 1996).

Table 6: Levels of overconfidence in each elicitation.

	Geography	History	Literature	Science
OE	8%	-1%	4%	5%
OPL	-14%	-7%	-16%	-19%
OPR	43%	24%	34%	41%
OE (alternate)	-17%	-22%	-23%	-23%

Note. We adopt the same nomenclature as before—OE is overestimation, OPL overplacement, OPR overprecision, and OE (alternate) is the new measure of overestimation. The domains are the columns.

We next move to testing the correlations among the different measures of overconfidence. The 16 x 16 correlation matrix is too large to include (and we include an analogous matrix in Study 1). The results are broadly the same. Of the 120 possible pairwise correlations, every single one was positive and statistically significant at an alpha of 0.05. The correlations within-elicitation across domain were strong for overestimation (.36 to .49), overplacement (.39 to .54), overprecision (.50 to .72), and with the alternate measure of overestimation (.41 to .52).

The correlations between the two measures of overestimation that were computed using the same data (e.g., overestimation in geography with overestimation (alternate) in geography) ranged from .69 to .77. The correlations between overestimation measures that neither shared the same method nor domain (e.g.,

overestimation in geography with overestimation (alternate) in history) had an average of .30. In general, the different measures of overestimation were highly correlated—the 8 measures had an overall alpha of .85.

To provide further clarity, we averaged each elicitation across domains, within-elicitation and include the correlations among these averages in Table 7. The correlations were all positive and significant, with magnitudes varying somewhat. In sum, we found strong evidence of positive associations across different elicitations and domains. Though the absolute levels of overconfidence observed varied, respondents' relative positions in each distribution were strongly associated.

Table 7: Correlations among each form of overconfidence.

	M	1	2	3
1. OE	4%			
2. OPL	-14%	0.47**		
3. OPR	36%	0.41**	0.35**	
4. OE (alternate)	-21%	0.70**	0.55**	0.24**

Note. We adopt the same nomenclature as before—OE is overestimation, OPL overplacement, OPR overprecision, and OE (alternate) is the new measure of overestimation.
Note. * indicates $p < .05$; ** indicates $p < .01$

3.2.5.2 Estimating core overconfidence

Having established that there are positive associations among different elicitations and measures of overconfidence, we next turned to estimating core overconfidence, as in Study 1. To do so, we standardized each of the 12 measures of

overconfidence (we do not use the alternate measure of overestimation going forward) within domain-elicitation, and entered them into a mixed effects model, with a fixed intercept and random intercepts nested within respondents. The random intercepts were our resulting measure of core overconfidence.

As before, we wanted to test the effectiveness of core overconfidence at predicting hold-out samples of overconfidence measures. To do so, we estimated overconfidence with 8/12 measures at a time, dropping one elicitation (e.g., overestimation), estimated core overconfidence with this subset of the data, and predicted each of the 4 dropped measures with the core overconfidence estimate. When dropping overestimation from the estimation of core overconfidence, core overconfidence correlated with each overestimation measure at .38 (geography), .45 (history), .37 (literature), and .41 (science, all *p*s < .001). The corresponding numbers for overplacement were also high (.65, .64, .49, .61), as were those for overprecision (.32, .41, .37, .36). All were significant at $p < .001$. In short, we again find that core overconfidence strongly predicts hold-out samples of overconfidence in a different domain. This suggests that there is a common element that is shared by different elicitations that contributes to the observed overconfidence in each case. As before, we computed the pooling coefficient (w) in these data. The value for 1 – w = 0.846, which was similar to Study 1. This suggests that our measurement of core overconfidence was highly reliable.

3.2.5.3 Confidence, accuracy, and overconfidence

Overconfidence captures people's calibration: are they more confident than is warranted, given their actual performance or ability? Yet the way it is measured also allows considering people's confidence, unconditional on their performance. Clarifying the relationships between confidence and overconfidence both within and across elicitations may help to clarify the nature of any dispositional component to the observed results. We investigate this question next.

We first transformed the confidence measures for overprecision such that they moved in the same direction as the other confidence measures. We did so by taking the log of the MEAD and multiplying it by minus one.[15] A higher score on the resulting variable indicated *greater* narrowing of one's confidence intervals (i.e., higher confidence). We performed the same transformation for the accuracy measure for overprecision—respondents' mean absolute deviation (MAD)—such that a higher score indicated greater accuracy. We then standardized each confidence and accuracy measure within domain-elicitation, as we did for the overconfidence measures when computing the correlations among them, before taking the average for each across domains. We tabulate the correlations among these measures in Table 8.

[15] These values were infinite for some observations; we conducted these analyses with complete sets of data (N = 492). This explains small discrepancies between the data and previously reported numbers.

Table 8: Correlations among overconfidence, confidence, and accuracy measures.

Form	Variable	1	2	3	4	5	6	7
Overestimation	1. OC							
	2. Confidence	.69** [.64, .73]						
	3. Accuracy	-.50** [-.56, -.43]	.23** [.15, .32]					
Overplacement	4. OC	.48** [.40, .54]	.33** [.24, .40]	-.27** [-.35, -.19]				
	5. Confidence	.48** [.41, .55]	.58** [.52, .64]	.01 [-.08, .10]	.69** [.64, .73]			
	6. Accuracy	-.02 [-.10, .07]	.29** [.21, .37]	.36** [.28, .43]	-.45** [-.51, -.37]	.34** [.26, .42]		
Overprecision	7. OC	.34** [.26, .42]	.07 [-.02, .15]	-.41** [-.48, -.34]	.35** [.27, .43]	.07 [-.02, .15]	-.37** [-.45, -.30]	
	8. Confidence	.22** [.14, .30]	.18** [.10, .27]	-.12** [-.20, -.03]	.06 [-.02, .15]	.10* [.01, .18]	.04 [-.05, .12]	.49** [.42, .56]
	9. Accuracy	-.31** [-.39, -.23]	.13** [.04, .21]	.58** [.52, .64]	-.39** [-.46, -.31]	-.01 [-.10, .08]	.49** [.42, .55]	-.69** [-.73, -.64]

Note. OC is overconfidence, and confidence and accuracy refer to their respective measures as defined in the Methods section of Study 2 and in the paragraph above.
Note. ** $p < .01$, * $p < .05$

First, we tested the correlations between the different measures of *confidence*. The confidence measures for overestimation and overplacement were significantly positively correlated ($r = 0.58$, $p < .001$). Confidence in the overprecision measure was also positively associated with confidence in the overestimation ($r = 0.18$, $p < .001$) and overplacement paradigms ($r = 0.09$, $p = .038$), though to a weaker degree. Further, each confidence measure significantly correlated with overconfidence in that elicitation; for overestimation ($r = 0.69$), overplacement ($r = 0.69$), and overprecision ($r = 0.49$), all were positively and statistically significant ($p < .001$).

Next, we consider the relationships among measures of accuracy. Accuracy was significantly correlated across overestimation and overplacement ($r = 0.36$, $p < .001$), among overestimation and overprecision ($r = 0.58$, $p < .001$), and among overprecision and overplacement ($r = 0.49$, $p < .001$). These positive associations may be produced by domain knowledge, as well as knowledge being correlated across domains. In each case, accuracy was negatively associated with overconfidence (overestimation: $r = -0.50$, overplacement: $r = -0.45$, overprecision: $r = -0.69$, all $p < .001$). Overall, both confidence and accuracy measures were related with overconfidence, and were related with other measures of the same type (i.e., confidence in overestimation was associated with confidence in overprecision).

We further consider the importance of these different elements to dispositional overconfidence. It could be the case that there is a "high confidence type" of person who consistently indicates high confidence across elicitations. The fact that confidence is correlated across elicitations and that confidence is related to overconfidence within each elicitation opens this possibility. This would be one form of miscalibration. Another is the following: Suppose that within an elicitation, for a given level of confidence some degree of overconfidence is expected. This could be estimated by regressing overconfidence on confidence. If someone exhibited *more overconfidence than was expected*, they would have a positive residual term in that regression model. Another test for an "overconfident type" of person could examine whether the residuals from such

regressions were correlated *across* the elicitations. This would show whether there was a type of respondent who consistently displayed more overconfidence than was expected, given their levels of confidence. Accordingly, we regressed each overconfidence measure on its respective confidence measure and stored the associated residuals. Among the 12 sets of residuals (from regressing each of the 12 overconfidence measures on each of the 12 confidence measures) we computed 66 pairwise correlations. Of these 66, 62 were positive and significant; the other 4 were null (and were all associated with overplacement in literature). Within an elicitation, the correlations among the residuals were quite strong; for overestimation they ranged from, $r = .18$ to $r = .38$, for overplacement from $r = .11$ to $r = .45$, and for overprecision from $r = .45$ to $r = .61$. We propose that this may suggest that the core overconfidence factor does not *merely* reflect confidence, but also further determinants above and beyond what might be expected for given levels of confidence.[16]

Overall, there were several key takeaways from these analyses. First, we found that confidence judgments were positively correlated across the elicitations. This

---

[16] Note that in this analysis, the correlations among these residuals could reflect accuracy. For a given level of confidence, there will be respondents with lower or higher objective performance. Without accounting for this variation in performance/accuracy, the residuals could reflect accuracy itself rather than calibration.

To address this question, we ran an additional analysis—we regressed each domain-level overconfidence measure on its confidence measure and the respondent's average accuracy across the 11 other tasks. We then extracted the residuals from these regressions and tested the pairwise correlations among them. After accounting for accuracy, we only observed consistent, positive correlations among the residuals from the overprecision measures. Across the six correlations from the overprecision measures, the average size was .34.

provides evidence for a subjective component that may contribute to each form of overconfidence, supporting earlier research (Larrick et al., 2007). However, we also found that people with above expected overconfidence for a given confidence level in a given elicitation were more likely to exhibit the same tendency in other elicitations. This provides evidence for a common factor underpinning measurements of overconfidence above and beyond confidence. In further tests, it appears that a combination of confidence and accuracy primarily explains overestimation and overplacement, whereas overprecision measures still exhibited correlated residuals when controlling for these variables, which could suggest the presence of a distinct, calibration factor.

3.2.5.4 An alternate measure of core overconfidence

As a construct, core overconfidence has desirable properties—it correlates with hold-out measures of overconfidence, individual differences, explains significant variation in individual domains and elicitations, and appears to be stable even over long time horizons. Yet it does require a large number of items to be measured reliably. We next investigate whether it is possible to construct a short 3-item test that can efficiently predict core overconfidence. To be useful, this measure would have to correlate highly with the long-form measure of core overconfidence (e.g., above $r = 0.5$). We attempt to construct such a measure by analyzing the data from our *overconfidence innovation* items. These included eight measures designed to assess core overconfidence in unison, and a ninth item that used the same logic as the "Single Item Narcissism Scale" (Konrath et al.,

2014), asking people how overconfident they are to see if they have metacognitive awareness of their own overconfidence.

Across the eight main *overconfidence innovation* items, there was one that required two dollar amount responses (OCI1), one that was a multiple choice question (OCI2), and six that required numeric entries (OCI3-8). The two dollar amount responses from OCI1 both strongly correlated with core overconfidence ($r = 0.52$ and $r = 0.50$, both $p < .001$), and the six numeric entries all positively correlated with core overconfidence (rs ranging from 0.17 to 0.60). In the multiple choice question, time spent working on your own product was negatively associated with overconfidence ($r = -0.12$, $p = .006$) whereas time spent working on 'Other' was positively associated ($r = 0.15$, $p = .001$).

We chose to construct an "Overconfidence Test" (OCT) from the three numeric entry items that correlated most strongly with core overconfidence.[17] These were OCI3, OCI5, and OCI7 (see the Methods section for the full items). We averaged people's responses to these three items to create an index measure, the OCT. The correlation between OCT and core overconfidence was $r = .63$ ($p < .001$). In other words, the OCT was able to explain significant variation in core overconfidence with just three items. The three items had an alpha of 0.65. Though this is below the thresholds used for

[17] We omitted the two dollar entry response from consideration due to a confound with risk aversion, and preferred not to include the multiple choice "Other" category because we did not have a theoretical rationale for this response's positive association with core overconfidence.

longer tests, the OCT appears to be useful as a short measure that taps into core overconfidence efficiently.

Finally, we also tested for the correlation between people's self-reported overconfidence (OCI9) and their actual core overconfidence. This correlation was .45 ($p <$ .001) which indicates that people have some awareness of how overconfident they are. This result is quite unexpected as overconfidence is generally seen as a calibration error—a failure of people to map their own knowledge into appropriate levels of confidence. If people are aware of this miscalibration, it raises questions about whether this characterization is appropriate.

3.2.5.5 Stability over time

We were also interested in testing the stability of core overconfidence and related constructs over time with a lower attrition rate than the long time-horizon callback in Study 1 (attrition was just 11% in Study 2, with 89% of respondents returning). To this end, respondents completed our measures of overconfidence twice, with a minimum of a one-week delay between the sessions. We computed core overconfidence and OCT as before—using mixed effects modelling for the former, and the average of three items (OCI3, OCI5, and OCI7) for the latter. Overestimation, overplacement, and overprecision were computed as the average of their four domains (which were standardized within domain-elicitation first). Below, we discuss the stability of core overconfidence and its constituent measures over this time period, as well as the stability of a short measure of

core overconfidence (OCT) and participants' self-reports of their overconfidence. These correlations are found in Table 9. Finally, we include a further validation of core overconfidence's stability using respondents' narcissism scores.

Table 9: Correlations among measures of overconfidence over two time periods.

	Time 2 measures					
Variable	Core OC	OE	OPL	OPR	OCT	Self-report
Core OC	.85** [.83, .88]	.65** [.60, .70]	.66** [.61, .71]	.70** [.65, .75]	.57** [.50, .63]	.34** [.25, .42]
OE	.65** [.60, .70]	.72** [.68, .77]	.42** [.33, .49]	.42** [.35, .50]	.51** [.44, .58]	.30** [.21, .38]
OPL	.65** [.59, .70]	.42** [.34, .50]	.76** [.72, .80]	.35** [.27, .43]	.50** [.43, .57]	.28** [.19, .36]
OPR	.67** [.62, .72]	.39** [.30, .46]	.37** [.28, .44]	.81** [.78, .84]	.31** [.23, .39]	.21** [.12, .30]
OCT	.54** [.47, .60]	.50** [.43, .57]	.49** [.42, .56]	.29** [.21, .38]	.68** [.63, .73]	.47** [.39, .54]
Self-report	.38** [.30, .46]	.39** [.31, .47]	.34** [.26, .42]	.18** [.09, .27]	.39** [.31, .47]	.58** [.52, .64]

Note. We adopt the same nomenclature as before—OE is overestimation, OPL overplacement, OPR overprecision, and OE (alternate) is the new measure of overestimation. The OCT is the average of the three OCI items (OCI3, OCI5, and OCI7). The self-report is responses to OCI8, where participants self-reported their own overconfidence.
Note. The way to read the rows and columns is the same as Tables 4 & 5: the correlation between time 1 OPL and time 2 core OC is $r = .65$; and the correlation between time 1 core OC and time 2 OPL is $r = .66$.
Note. * $p < .05$, ** $p < .01$

We found that core overconfidence was highly correlated over time ($r = 0.85$, $p < .001$), corroborating the results of the first study. This provides further strong evidence that there is a dispositional component to overconfidence. If the correlations among different measures of overconfidence observed across Studies 1-2 were purely driven by situational factors, we would not expect them to exhibit significant stability over time (both over 1-week and 2.5 years). Study 2 provides an important complement to the other call back study as it had a much lower attrition rate (11%)—showing across the majority of the initial sample that the temporal stability was robust.

We were also able to test the stability of the constituent measures of core overconfidence—for example, was overestimation more stable than overprecision? We observed similar (and very strong) correlations between the time 1 and time 2 measures of each of the elicitations; for overestimation ($r = 0.72$, $p < .001$), overplacement ($r = 0.76$, $p < .001$), and overprecision ($r = 0.81$, $p < .001$), all were strong. These were relatively similar to the results of Study 1, in which overprecision was the most stable, followed by overestimation, and then overplacement. Overall, it appears that overprecision is the most stable form of overconfidence over time, which could be due to its weaker relationship with people's subjective views of themselves.

Further, we found that overestimation at time 1 significantly predicted overplacement in time 2 ($r = 0.42$, $p < .001$) and other cross-elicitation results that can be seen in Table 9. In these correlations, both time period and elicitation varied, reducing

the effects of momentary situational influences (e.g., mood) and elicitation-specific influences, isolating more clearly a common, dispositional component. These correlations were generally quite strong, ranging from .35 to .42, with a median of .41. Participants' responses to the OCT also exhibited stability over time, suggesting that this efficient measure of core overconfidence also taps into a dispositional component of overconfidence ($r = 0.68$, $p < .001$). Finally, even the single item measure *asking* people how overconfident they are indicated substantial stability ($r = 0.58$, $p < .001$). The lower stability of the self-reports may reflect their closer relationship with situational factors (e.g., a subjective *feeling* of how overconfident one was at the time).

We include an additional validation of the stability of core overconfidence over time, which we describe next. From Study 1, we know that narcissism (as captured by the Narcissistic Personality Inventory (Ames et al., 2006)) is strongly associated with core overconfidence. In Study 2, we included this measure only in the second stage. We tested whether, when measured in the second stage, it had a positive correlation with core overconfidence in the first stage of the study. The rationale behind this test is the following: if there is a strong situational component to overconfidence that permeates through all measures, people answering the *same* items in stage two may recall how they felt in the situation of stage one, and so this situational effect could contaminate the data across the two sessions. Given that narcissism was not measured in the first stage, this problem is not shared by the correlation between stage one core overconfidence and

stage two narcissism. The positive correlation observed in this relationship ($r = 0.50$, $p < .001$) provides further evidence that there is a dispositional component to core overconfidence.

There is another piece of evidence that results from the inclusion of narcissism as a variable: the difference between the size of correlation between narcissism and core overconfidence in the *same* time period, and narcissism and core overconfidence in different time periods. In Study 2, narcissism measured at stage two correlated with stage two core overconfidence at $r = .49$, and with stage one core overconfidence at $r = .50$ (actually *increasing*, though very slightly). In Study 1, narcissism at time one correlated with time one core overconfidence at $r = .58$ and with time two core overconfidence at $r = .50$; and narcissism at time two correlated with time two core overconfidence at $r = .41$ and with time one core overconfidence at $r = .38$. All this to say, narcissism exhibited a similar strength of relationship with core overconfidence *measured in a different time period* to the strength observed when both were measured in the same period. This further supports our claim that there is a dispositional component to core overconfidence.

3.3 Discussion

In Study 2, we again found, on average, a small degree of overestimation, a larger degree of underplacement, and very strong pattern of overprecision. The pairwise correlations between *every* measure of overconfidence were positive and significant

(with very strong correlations across domain, within elicitation). Our measure of core overconfidence predicted out-of-sample measures of overconfidence, and our overconfidence measurements exhibited significant reliability. These results all replicated those of Study 1.

In addition to these results, we further probed the role of *confidence* in our data. We found that there is evidence for both a confidence factor—people who feel confident in one task are likely to feel relatively more confident in another task—but also an *overconfidence* factor, where people who exhibit higher overconfidence than they "should" for a given level of confidence tend to do so across tasks. In addition to these results, we constructed a 3-item "OCT" measure that correlated with the long form core overconfidence at $r = .63$, and found that people appear to have meta-cognitive awareness of their own overconfidence.

Finally, we expanded on our analysis of the stability of overconfidence *over time* in a larger sample with a low attrition rate. Respondents exhibited remarkable temporal stability in both their core overconfidence, its constituent measures, and their responses to the OCT items. All of this points towards a dispositional component to overconfident beliefs. Across short and long time horizons, core overconfidence was stable.

One notable limitation of the study is that we used a convenience sample on Amazon's Mechanical Turk. It is important to replicate these results in a new sample, and in a sample that is more representative of the U.S. population. For example, we

observed a very high correlation between narcissism and overconfidence across Studies 1-2. If the narcissism observed in the samples was abnormally high or consistent, this could have generated correlations between measures of overconfidence that may not be present in a new sample.

In Appendix B, we report the results of a study using a Nationally representative sample which similarly finds positive correlations among measures of overconfidence and correlations between core overconfidence and different traits. This study found that both self-view and cognitive style related traits were associated with core overconfidence. Decomposing core overconfidence back into its constituent measures (overestimation, overplacement, overprecision) suggests that there may be two clusters of traits that are associated with core overconfidence: self-view variables that determine overestimation and overplacement, and cognitive style variables that determine overprecision. That said, some of the strongest individual difference correlates were associated with all three forms (e.g., actively open-minded thinking, (Baron, 2019)). These results are in line with our theorizing that the strong subjective component of the overestimation and overplacement elicitations that captures people's self-views would be strongly associated with such variables. We do not further emphasize this study as we reduced the number of overconfidence measures to 3, and this severely affected the reliability of the resulting measurement of overconfidence.

In sum, in Study 2 we found further support for the results of Study 1, along with new results relating to the role of confidence, the ability to measure core overconfidence more efficiently, and the stability of different measures of overconfidence over time. A further supplemental study highlighted the robustness of positive associations among measures of overconfidence to design choices and sample populations, but highlighted the need for reliable measurement of core overconfidence using a wide array of questions.

Chapter 4: General Discussion

In the first chapter, we began by reviewing literature regarding the different forms of overconfidence (overestimation, overplacement, and overprecision) and how they are related to each other. There are reasons to believe that these forms might be empirically distinct (e.g., the different levels of overconfidence observed, and the differential effect of task difficulty across overestimation and overplacement), but also reasons to believe there is a dispositional component to overconfidence (e.g., different forms of stability of overconfidence measures, and overconfidence's relationship with measures of stable traits). A dispositional component to overconfidence would imply that the different forms should be related. In the present research, we aimed to further understand the relationships among the different forms of overconfidence, the nature of overconfident beliefs, and the relationship between overconfidence and individual difference measures. We discuss our findings below.

4.1 Summary and contributions

4.1.1 The existence of a dispositional component of overconfidence

Our primary claim is that there is a dispositional component to overconfidence, as well as situational and elicitation effects. In our data, we did observe different average levels of overconfidence across elicitations, with there being substantial overprecision, some overestimation, and mostly underplacement. There were also domain-elicitation interactions (i.e., moving from an easier domain to a harder one had different effects on

the average overconfidence displayed in different elicitations). These results are consistent with findings in the literature (e.g., Klayman et al. 1999; Moore & Healy, 2008; Juslin et al., 1999; Erev et al,, 1994; Burson et al., 2006; Larrick et al., 2007; Kruger, 1999; Moore & Small, 2007; Moore & Healy, 2008), and show that elicitations do indeed have effects on the observed level of overconfidence. However, this is not incongruous with there being a dispositional component to overconfidence that is apparent when we observe the same people's behavior across tasks.

We found significant positive correlations between *every* overconfidence measure across domains and elicitations in our three studies. These correlations suggest that a person who displays greater overestimation is likely to also display greater overplacement and overprecision, and so on. This is evidence of an association across different forms of overconfidence at the level of the individual, suggesting that a common element does underpin the different elicitations.

These correlations held when all of the overconfidence measures were taken at the same point in time and when measured across time. We retrospectively added a second wave of data in Study 1 and prospectively created two waves in Study 2. This helped us to identify the stability that exists above and beyond the role of momentary situational factors that could inflate individual level correlations at one point in time. For example, if those experiencing positive affect indicated greater overconfidence across tasks, this could cause people's overconfidence in the tasks to be correlated without there being a dispositional factor. To assess this possibility, we needed to confirm the stability of overconfidence over time.

Having looked at the relationships among different elicitations, we turned our attention to *core overconfidence*. We measured this by standardizing respondents' 12 measures of overconfidence within domain-elicitation, entering these into a mixed effects model with random

intercepts, and extracting these random intercepts as our measure. This core overconfidence is an average across domains and elicitations, and aims to reduce our measurement's dependence on the idiosyncrasies of the tasks. In two studies we tested the test-retest reliability of this core overconfidence, both over a 2.5 year time period, and a 1-week time period. In both cases we found remarkable intertemporal stability (r = .77 and r = .85), suggesting that there is a dispositional component to the observed overconfidence.

Overall, we both replicated some existing elicitation effects (such as overconfidence being stronger when measured as overprecision than as overestimation) and provided strong evidence that there is a dispositional component to overconfident beliefs. In other words, the understanding that we proposed in Chapter 1, Section V is supported. These findings contribute to the literature in that they highlight the consistent and important role that *dispositions* play in determining overconfidence in tasks, in unison with other factors. Our results are not incongruous with existing findings about the effect of elicitations and task difficulty on *average* levels of overconfidence, but rather a complement to them. We find significant evidence that someone who is *relatively* more overconfident in one task is likely to be *relatively* more overconfident in another task. This perspective is a departure from claims that the different forms of overconfidence are empirically distinct (Moore & Healy, 2008).

4.1.2 What is the nature of this dispositional component?

Having found evidence that there is a dispositional component to overconfidence, we next asked, is this factor "just" confidence? It could be the case that people walk around with dispositionally different levels of confidence, and these repeatedly show up in measures of overconfidence. This would be less about *overconfidence* (the calibration between knowledge and confidence judgments), and more about confidence. Though we did find that confidence was also related across tasks, we also found clear evidence for another factor, which may reflect

knowledge or calibration (or both). Those who exhibited overconfidence above that which was expected for a given level of confidence were likely to do so across tasks.

In other words, we found evidence that people may have dispositionally worse calibration of their confidence judgments to their actual performance or ability. This suggests the associations across different forms of overconfidence are driven by more than just a common subjective component of how *good* people feel (Larrick et al., 2007), but also how the mapping between one's subjective experiences and reality. Further work is required to tease apart what proportion of the stability in overconfidence can be attributed to confidence versus other factors.

In an additional test, we found that people's self-reports of how overconfident they are had a strong relationship with their actual core overconfidence. This suggests that people have meta-cognitive awareness of their overconfidence. Given that the literature has often portrayed overconfidence as a failure in metacognition, this finding is somewhat surprising and deserves further exploration.

4.1.3 Core overconfidence and individual differences

Our other research questions concerned the relationship between core overconfidence and other individual difference measures (i.e., traits). Past studies have tended to use a single elicitation method which may have rendered relationships susceptible to the idiosyncrasies of the elicitation. By using core overconfidence as our measure, we aimed to establish relationships between the dispositional component of overconfidence and other traits. We found very strong relationships between core overconfidence and traits in Study 1 (especially with narcissism, corroborating earlier research (Campbell et al., 2004)). Yet even there, the correlations did not approach 1. This reflects the fact that overconfidence is not *only* dispositional, but rather does contain dispositional and situational components, along with elicitation effects. As a result, the relatively more modest correlations are to be expected.

One concern that might arise from these findings is that the *stability* in core overconfidence is driven by another individual difference (e.g., narcissism) and that the role of such a variable is observed despite the situational effects, rather than constituting a distinct, dispositional component of overconfidence. Our data allow us to gather some evidence related to this hypothesis. Using Study 1's data, we regressed core overconfidence at time 2 on both time 2 narcissism and time 1 core overconfidence. We found a positive significant effect of lagged core overconfidence ($b = 0.639$, $p < .001$) above and beyond contemporaneous narcissism ($b = 0.302$, $p = .069$), suggesting that the intertemporal stability of overconfidence is at least partially explained by unique, stable differences in overconfidence. The same test in Study 2's data with a shorter time delay found similar results; lagged core overconfidence had a significant positive effect ($b = 0.858$, $p < .001$) even when controlling for contemporaneous narcissism ($b = 0.265$, $p < .001$).

4.2 Limitations

It is a limitation of the present research that we are not able to further distinguish between the components of a measure of overconfidence other than to provide evidence of their existence. For example, in a given measure of overconfidence, what is the contribution of dispositional, situational, and elicitation effects? Further research is required to discern the relative importance of these components, which would also help to better interpret results in the literature. That said, there is also a pragmatic approach to considering the importance of the components. For example, consider a manager who is deciding between different members of their team in order to select one to complete a work task. They want to measure each team member's overconfidence in order to

predict the overconfidence that they will display in the work task. If the work task is to be completed immediately after, situational factors such as mood are largely held constant. If the work task is of a similar nature to the measure of overconfidence (i.e., its response format is similar), elicitation effects may be minimized. The dispositional component will of course be more constant—but the importance of its contribution will be determined by variation in the other parts. In other words, measuring core overconfidence may be useful in practical settings even if it is not possible to fully extricate the contributions of its different components to its measurement.

Another limitation of our consideration of measures of overconfidence is our limited specification of the nature of each elicitation's idiosyncrasies. For example, if you consider overestimation as a baseline, overplacement differs in that it asks for comparison to a reference group, and overprecision asks for multiple quantiles of a distribution, but this comparison misses many other elements. For example, overestimation in our studies uses binary choices, whereas overplacement asks for single numeric estimates. One limitation of the present work is summarized by the question: How should we think about these idiosyncrasies? They appear to be distinct from the beliefs that underpin each measure (which are driven by dispositional and situational factors), but it is impossible to measure overconfidence without invoking some set of these parameters. These idiosyncrasies also vary in consequence. For example, asking for a 90% confidence interval will produce more overprecision than a

50% confidence interval without any difference in the underlying beliefs, but the comparative element of overplacement may engage distinct psychological processes that do alter beliefs. In the present research, we have aimed to "average out" these effects by always asking multiple elicitations and domains, but their role is nebulous, and could be important in unobserved ways.

A key caveat to our results is that they capture associations among (mostly standardized) variables. In other words, we test the correlations between different variables, such as overestimation in geography and overprecision in history. This is a *rank order hypothesis*. A positive correlation among the two does not tell us about the levels we will observe in each case. It could be the case that the two were positively correlated, but all datapoints indicated *negative* overconfidence (i.e., underconfidence). In this sense, our primary analyses allow us to answer one set of questions (e.g., If someone exhibits relatively more overestimation, are they also likely to exhibit relatively more overplacement?) but not another set of questions (e.g., If I observed someone's positive overestimation score, are they likely to also exhibit positive overplacement?) The focus on different questions could explain some of the disparities between our conclusions and extant research (e.g., Moore & Healy, 2008). However, by converting our data back to its original response scales, we are able to provide some evidence on such questions. If we use the full data from Stage 1 of Study 2 (N = 494) and look at simple regression models containing an intercept and a single form of overconfidence as

a predictor, we can compute the level of one form of overconfidence that would lead to us predicting a *positive* overconfidence value in a different measure.

Predict positive overestimation when…	Overplacement above -30%
Predict positive overestimation when…	Overprecision above 19%
Predict positive overplacement when…	Overestimation above 20%
Predict positive overplacement when…	Overprecision above 73%
Predict positive overprecision when…	Overestimation above -49%
Predict positive overprecision when…	Overplacement is any value

These predictions help to provide some intuition for the nature of the relationships in Study 2's data—for example, given that overplacement has a negative mean value (i.e., people on average underplace), it takes large degrees of overconfidence in the other forms for the regression models to predict *overplacement*. It is important to note that this is not a *general* finding, but rather a way of communicating the effect sizes in our data. The choice of elicitations, domains, task difficulties, and many other variables will all contribute to such thresholds. These analyses also raise another point—if overconfidence as a construct is *only* values above 0, rather than a full range between -100% (underconfidence) and 100%, then this has implications for the interpretation of our data. Yet if we filter Study 2's data to just include participants who were

overconfident in each of the three forms (overestimation, overplacement, and overprecision N = 89), we still observe strong positive correlations among the different forms (*r*s between .43 and .61, all $p < .001$).

Another way of connecting our effects to the broader literature is translating the different levels of core overconfidence into predicted levels of overconfidence for each form on the original response scales. Using the data from Study 2 across both stages (N = 442) we plotted Figure 2. Along the x-axes are either core overconfidence (left-hand panel) or OCT score (right-hand panel) in the first time period of the study. Along the y-axis is predicted overconfidence at the second time period for a given level of core overconfidence / OCT, on its original response scale. There are a few striking takeaways: at 1 SD below mean levels of core overconfidence, the predicted overconfidence is essentially all negative (-15% underestimation, -46% underplacement, 0% overprecision), but at 1 SD above mean core overconfidence, all are positive (21% overestimation, 24% overplacement, and 68% overprecision). Another point to note is that overplacement exhibits a steep slope despite its lower overall levels, and that overestimation has a shallower slope than the other two. This could be due to the 50% expected accuracy rate in the 2AQ paradigm leading to reduced responsiveness of the observed overestimation. As we turn to the second panel, there is a similar pattern of results. OCT ($M = 30.6$, $SD = 14.7$) at time one predicts all three forms of overconfidence at time two, and overplacement and overestimation cross over at higher levels of OCT.

Overall, though our primary hypotheses concern the relationships between standardized variables, converting these data back into their original response scales provides results consistent with our assertions.

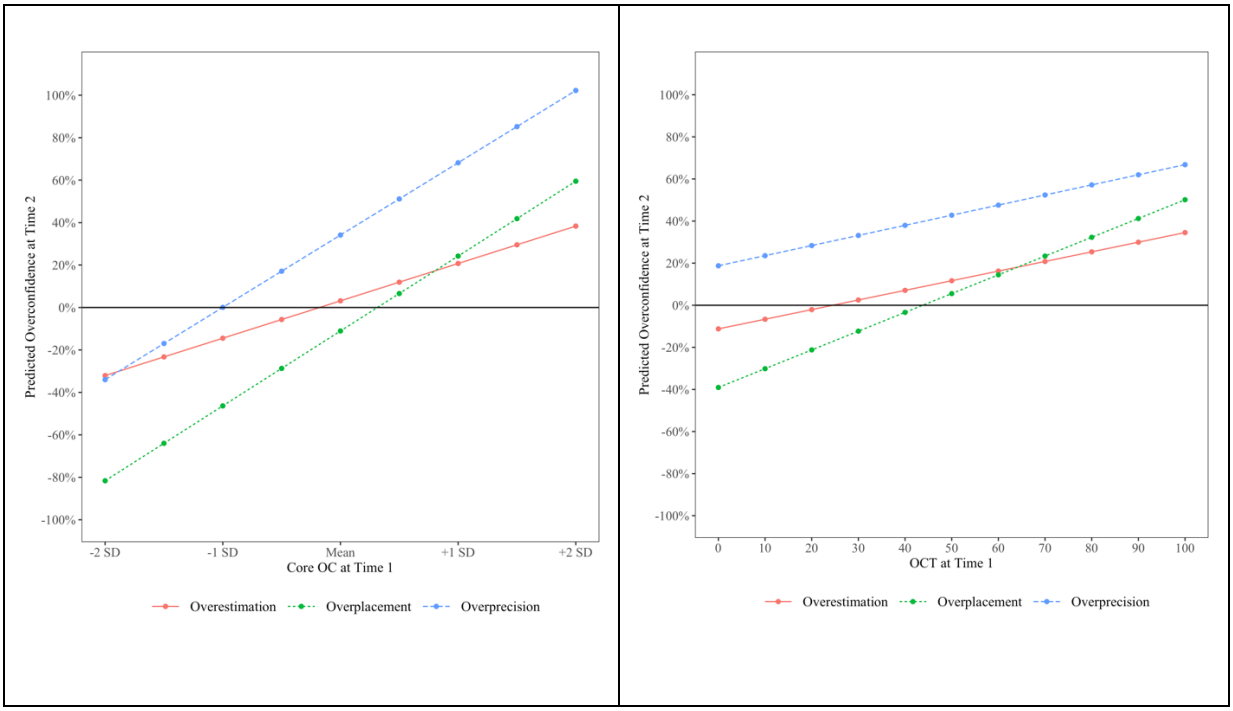

Figure 2. Predicted overconfidence on original response scale for different levels of core overconfidence and OCT.

Three other limitations are the long number of questions required to reliably measure core overconfidence, the domains and items we used, and the convenience samples used in Studies 1-2. In our main studies, people answered long strings of questions that involve both problem solving / general knowledge, and confidence judgments. The overconfidence part of our surveys can take 30-40 minutes. When we attempted to shorten this length in the supplemental study, the reliability of our measurement was greatly diminished (though this may have also been due to the

change in sample population). It could be the case that when people complete long tasks they settle into habits or use heuristics that are not representative of their behavior in shorter tasks (in this vein, Klayman et al. (1999) also hypothesized that people might exhibit less overconfidence in larger samples of questions, as they became more calibrated in their processes, though no strong effect was found). Further investigation is required into whether the decision processes people exhibit in long experimental sessions are predictive of their performance in shorter sessions.

As far as the domains and items are concerned, though we attempted to broaden the generality of our conclusions by using multiple domains, and to provide representative numbers of contrary questions by randomly sampling from domains (per Gigerenzer et al., 1991; Klayman et al., 1999), our specific materials may still have been relevant. Finally, we note that the convenience samples from MTurk may not be representative of the broader population. Though we did replicate the positive correlations among different measures of overconfidence in a supplemental study, the magnitude of these relationships was attenuated. Further work is required to establish what proportion of the discrepancy between Studies 1-2 and the supplemental study is attributable to reduced reliability, random chance, and systematic variance in study populations.

4.3 Future directions

The present research provides compelling evidence that there is a dispositional component to overconfidence. This opens many possibilities for future research. The most immediate prospect is further investigating how to measure core overconfidence more efficiently. As noted in the limitations, the measurement of core overconfidence in Studies 1-2 is onerous for respondents. This may preclude its use in organizational settings and shorter experiments. Study 2 found initial evidence that a short 3-item "OCT" could meaningfully predict core overconfidence, but this should be further validated. The strong correlations between core overconfidence and traits (e.g., narcissism) in Study 1 also open the possibility that scale items could be borrowed from existing individual difference measures to create a "Frankenstein" scale to predict core overconfidence. Such an approach would further verify that the measure was capturing the dispositional component of core overconfidence, rather than idiosyncrasies of a response scale (as the numeric estimates of the OCT are a similar format to those indicated for our overestimation questions).

Another future direction concerns understanding more fully the role of both self-view and cognitive style variables in relation to core overconfidence. Though Study 1 suggested that self-views might be stronger correlates, we engaged in a more comprehensive investigation in a supplemental study in Appendix B and found that both types of trait played important roles. We found some evidence that there are two clusters of traits that are associated with core overconfidence: self-view variables that primarily inflated overestimation and

overplacement, and cognitive style variables that primarily reduced overprecision. This is in line with research suggesting that these elicitations (overestimation and overplacement) share a strong subjective component that is to do with how good a respondent feels when completing a task (Larrick et al., 2007). Further work is required to clarify the nature of overconfidence as an individual difference. Our data suggest that it is neither "just" a cognitive or self-view variable, but rather a construct that is related to both kinds of traits. Understanding the exact traits that are associated with core overconfidence could help to understand how biased search processes produce overconfidence effects, and the different components that contribute to holding overconfident beliefs.

Yet the possibility that core overconfidence is not unidimensional does raise some additional questions. For example, is it possible to distinguish between the different dispositional components of elicitations (e.g., overestimation and overprecision) that contribute to core overconfidence and the idiosyncrasies that influence the measurements in elicitations independently of overconfident beliefs? Resolving such questions is important in interpreting results that correlate only one form of overconfidence with another variable.

The existence of a dispositional component to overconfidence invites the possibility of asking how this component relates to people's behaviors. Two relevant areas are the relationship between overconfidence and risk taking, and the relationship between overconfidence and advice taking. In existing research, Malmendier and Tate (2005a) measure CEO overconfidence with two different kinds of measures, based on i) the CEO holding stock options longer than necessary, and ii) the language used to describe the CEO by the press, and using these measures find that overconfidence is related to a wide range of firm behaviors, including risk taking. For example, the sensitivity of investment to cash flows

(Malmendier & Tate, 2005a, 2005b), acquisition activity (Brown & Sarma, 2007), value-destroying merger activity (Malmendier & Tate, 2004), the likelihood of issuing equity, and the likelihood of using external capital markets (Malmendier, Tate & Yan, 2011). This literature seems to suggest that overconfidence is associated with increased risk taking. However, overconfidence is not defined *independently* from risk taking, as CEOs holding more stock in their own companies constitutes a greater financial risk than a more diverse portfolio. Measuring overconfidence independently of risk preferences could help to further clarify the nature of this relationship.

There is some existing field work that cleanly measures overconfident beliefs—Ben-David, Graham, and Harvey (2007, 2013) measure overconfidence by comparing executives' predicted probability distributions of stock market returns to the actualized rates of return. Ahmed and Duellman (2013) show using this measure that overconfidence is negatively associated with accounting conservatism—a concerning behavioral outcome. Yet by measuring core overconfidence, rather than a specific elicitation, we might hope to further understand the relationship between the dispositional component of overconfident beliefs and people's risk taking.

Theoretically, there are at least two ways in which overconfidence could be related to risk taking. It could be the case that more overconfident people are less risk averse, and so in choosing between gambles would show a greater preference for risky options. It could also be that more overconfident people only take greater risks when there is at least the *illusion* of a skill component to a game. An inflated view of one's abilities or reduced perceptions of uncertainty could lead to perceiving risks as subjectively smaller. This could induce greater risk taking for a given level of risk aversion. Relatedly, the relationship between overconfidence and risk taking may depend on whether the inherent uncertainty is epistemic or aleatory (Tannenbaum et al.,

2016). If the uncertainty is epistemic, overconfident people may be more prone to taking risks (excessive certainty in one's knowledge or the range of possible outcomes is a definitional feature of overprecision), whereas the prediction for aleatory uncertainty is less clear.

Another observation that one might make about the "overconfident" people in our lives is that they seem resistant to outside advice. However, little is known about this in an experimental context. Do overconfident people place less weight on the advice of others? It is possible to test this empirically. A traditional advice taking paradigm evaluates the *weight on advice* (WOA) by asking a judge for an initial estimate, providing them with advice, and asking for a final estimate (e.g., Soll & Larrick, 2009). The WOA is then calculated as the movement of the respondent's final estimate towards the advice. It seems likely that people who have higher dispositional overconfidence will weight outside advice less, due to excessive faith in their own estimates. This is an empirical question.

In an extension of this, more recent approaches to measuring advice taking have emphasized looking at the *influence of advice* (IOA)—the shift in a judge's subjective probability distribution over outcomes in response to advice (Soll et al., 2021)—rather than just the change in point estimates. Respondents are asked for an initial estimate as well as their confidence that the true answer lies within X units of their estimate. After receiving a point estimate of advice, respondents revise both their estimate and their confidence levels. This paradigm facilitates asking several additional questions: will core overconfidence predict initial confidence levels? How about the *change* in confidence in response to advice? These are open questions that are important future directions when considering the importance of the dispositional component to overconfident beliefs.

Chapter 5: Conclusions

Our goal was to investigate the nature of overconfident beliefs. Across studies, we found strong evidence that different measures of overconfidence are positively associated, spanning across elicitations and domains. These associations suggest that the measures share a common element. We found that core overconfidence exhibited remarkable stability over time (both 1-week and 2.5 years), suggesting that this common element contains at least a dispositional component. This conclusion is a relatively big departure from current consensus in the literature.

Some of the disparity between our conclusions and those of others result from our approach: We consider the individual as the unit of analysis, and pose hypotheses about people's *relative* overconfidence. Our main takeaway is that someone exhibiting *relatively* higher overestimation is likely to also display *relatively* higher overplacement (and so on for the other elicitation pairs). Yet such a person could actually be *underconfident* in both elicitations. If one considers task-level overconfidence or only overconfidence (not underconfidence) as the effects of interest, it is possible to reach different conclusions.

In this spirit, we hope that our perspective helps to highlight a different way of considering the nature of overconfident beliefs that emphasizes the importance of individual differences and opens the door to testing the relationship between the dispositional component of overconfident beliefs and various different outcomes.

Appendix A: Scales and measures

Narcissistic personality inventory (Ames et al., 2006)

Read each pair of statements below and select the one that comes closest to describing your feelings and beliefs about yourself. You may feel that neither statement describes you well, but pick the one that comes closest.

Item 1
I really like to be the center of attention. (Y)
It makes me uncomfortable to be the center of attention.

Item 2
I am no better or no worse than most people.
I think I am a special person. (Y)

Item 3
Everybody likes to hear my stories. (Y)
Sometimes I tell good stories.

Item 4
I usually get the respect that I deserve.
I insist upon getting the respect that is due me. (Y)

Item 5
I don't mind following orders.
I like having authority over people. (Y)

Item 6
I am going to be a great person. (Y)
I hope I am going to be successful.

Item 7
People sometimes believe what I tell them.

I can make anybody believe anything I want them to. (Y)

Item 8
I expect a great deal from other people. (Y)
I like to do things for other people.

Item 9
I like to be the center of attention. (Y)
I prefer to blend in with the crowd.

Item 10
I am much like everybody else.
I am an extraordinary person. (Y)

Item 11
I always know what I am doing. (Y)
Sometimes I am not sure of what I am doing.

Item 12
I don't like it when I find myself manipulating people.
I find it easy to manipulate people. (Y)

Item 13
Being an authority doesn't mean that much to me.
People always seem to recognize my authority. (Y)

Item 14
I know that I am good because everybody keeps telling me so. (Y)
When people compliment me I sometimes get embarrassed.

Item 15
I try not to be a show off.
I am apt to show off if I get the chance. (Y)

Item 16
I am more capable than other people. (Y)
There is a lot that I can learn from other people.

Coding

Participants get 1 for each item they indicate the narcissistic choice for; these are then summed together; their total score is out of 16.

Intellectual humility scale (Leary et al., 2017)

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.

Coding

All items responded to with Likert scales from 1 (Not at all like me) to 5 (Very much like me). Total score an average of responses to the 6 items.

Need for closure decisiveness facet (Kruglanski et al., 1993)

1. I would describe myself as indecisive. (R)
2. When I go shopping, I have difficulty deciding exactly what it is I want. (R)
3. When faced with a problem I usually see the one best solution very quickly.
4. I tend to put off making important decisions until the last possible moment. (R)
5. I usually make important decisions quickly and confidently.
6. Please check four if you are paying attention.
7. I tend to struggle with most decisions. (R)
8. When trying to solve a problem I often see so many possible options that it's confusing. (R)

Coding

All items responded to with Likert scales from 1 (Not at all like me) to 5 (Very much like me). There are 7 actual items (item 6 is an attention check). Of these 7 items, 5 were reverse coded. After reversing these 5, the total score is the average of the 7 items. A higher score indicates greater decisiveness.

The dominance facet of extraversion (Wiggins et al., 1988)

Please rate the extent to which the following words describe yourself.

How accurately does this describe you?

1. Self-assured
2. Self-confident
3. Assertive
4. Persistent
5. Firm
6. Dominant
7. Forceful
8. Domineering

Coding

All items responded to on Likert scales from 1 (Extremely Inaccurate) to 8 (Extremely Accurate). Total score an average of responses to the 8 items.

Cognitive Reflection Test (Frederick, 2005)

1. 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? (in cents)

___

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (in minutes)

___

3. 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? (in days)

___

Coding

Each item has a correct answer (5, 5, 47). Participants receive a score from 0 to 3 for how many items are answered correctly.

Berlin Numeracy Test (BNT; Cokely et al., 2012)

1. Imagine we are throwing a five-sided die 50 times. The sides are numbered 1 to 5. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

- 5 out of 50 throws
- 25 out of 50 throws

- 30 out of 50 throws

- None of the above

2. Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.

- 10%

- 25%

- 40%

- None of the above

3. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?

- 20 out of 70 throws

- 23 out of 70 throws

- 35 out of 70 throws

- None of the above

4. In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?

- 4%
- 20%
- 50%
- None of the above

Coding

Each item has a correct answer (3, 2, 1, 3). Participants receive a score from 0 to 4 for how many items are answered correctly.

Need for closure closed mindedness facet (Kruglanski et al., 1993)

Please answer the questions according to the scale provided.

1. Even after I've made up my mind about something, I am always eager to consider a different opinion. (R)
2. I dislike questions which could be answered in many different ways.
3. I feel irritated when one person disagrees with what everyone else in a group believes.
4. When considering most conflict situations, I can usually see how both sides could be right. (R)
5. When thinking about a problem, I consider as many different opinions on the issue as possible. (R)
6. I prefer interacting with people whose opinions are very different from my own. (R)

7. I always see many possible solutions to problems I face. (R)
8. I do not usually consult many different options before forming my own view.

Coding

All items responded to with Likert scales from 1 (Not at all like me) to 5 (Very much like me). There are 8 items. Of these 8 items, 5 were reverse coded. After reversing these 5, the total score is the average of the 8 items. A higher score indicates greater closed mindedness.

Optimism (Scheier et al., 1994)

Please answer the questions according to the scale provided.

1. In uncertain times, I usually expect the best.
2. It's easy for me to relax. (DECOY)
3. If something can go wrong for me, it will. (R)
4. I'm always optimistic about my future.
5. I enjoy my friends a lot. (DECOY)
6. It's important for me to keep busy. (DECOY)
7. I hardly ever expect things to go my way. (R)
8. I don't get upset too easily. (DECOY)
9. I rarely count on good things happening to me. (R)
10. Overall, I expect more good things to happen to me than bad.

Coding

All items responded to with Likert scales from 1 (Strongly disagree) to 5 (Strongly agree). There are 6 actual items (4 of them are meant to mitigate demand effects). Of these 6 items, 3 were reverse coded. After reversing these 3, the total score is the average of the 6 items. A higher score indicates greater optimism.

Self-efficacy (Chen et al., 2001)

Please answer the questions according to the scale provided.

1. I will be able to achieve most of the goals that I have set for myself.
2. When facing difficult tasks, I am certain that I will accomplish them.
3. In general, I think that I can obtain outcomes that are important to me.
4. I believe I can succeed at most any endeavor to which I set my mind.
5. I will be able to successfully overcome many challenges.
6. I am confident that I can perform effectively on many different tasks.
7. Compared to other people, I can do most tasks very well.
8. Even when things are tough, I can perform quite well.

Coding

All items responded to with Likert scales from 1 (Strongly disagree) to 5 (Strongly agree). Total score an average of responses to the 8 items.

The desire for control (Burger & Cooper, 1979)

Below you will find a series of statements. Please read each statement carefully and respond to it by expressing the extent to which you believe the statement applies to you.

For all items a response from one to seven is required. Use the number that best reflects your belief when the scale is defined as follows:

1. The statement doesn't apply to me at all.

7. The statement always applies to me.

Statements

1. I prefer a job where I have a lot of control over what I do and when I do it.
2. I enjoy political participation because I want to have as much of a say in running government as possible.
3. I try to avoid situations where someone else tells me what to do.
4. I would prefer to be a leader rather than a follower.
5. I enjoy being able to influence the actions of others.
6. I am careful to check everything on an automobile before I leave for a long trip.
7. Others usually know what is best for me. (R)
8. I enjoy making my own decisions.
9. I enjoy having control over my own destiny.
10. I would rather someone else took over the leadership role when I'm involved in a group project. (R)
11. I consider myself to be generally more capable of handling situations than others are.
12. I'd rather run my own business and make my own mistakes than listen to someone else's orders.
13. I like to get a good idea of what a job is all about before I begin.
14. When I see a problem I prefer to do something about it rather than sit by and let it continue.
15. When it comes to orders, I would rather give them than receive them.
16. I wish I could push many of life's daily decisions off on someone else. (R)
17. When driving, I try to avoid putting myself in a situation where I could be hurt by someone else's mistake.
18. I prefer to avoid situations where someone else has to tell me what it is I should be doing.
19. There are many situations in which I would prefer only one choice rather than having to make a decision. (R)
20. I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered by it. (R)

Coding

All items responded to with Likert scales from 1 (Not at all) to 7 (Always applies).

There are 20 items. Of these 20 items, 5 were reverse coded. After reversing these 5, the total score is the average of the 20 items. A higher score indicates greater desire for control.

Short Big Five measure (Gosling, Rentfrow, & Swann, 2003)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other

I see myself as:

1. Extraverted, enthusiastic.
2. Critical, quarrelsome. (R)
3. Dependable, self-disciplined.
4. Anxious, easily upset. (R)
5. Open to new experiences, complex.
6. Reserved, quiet. (R)
7. Sympathetic, warm.
8. Disorganized, careless. (R)
9. Calm, emotionally stable.
10. Conventional, uncreative. (R)

Coding

All items responded to with Likert scales from 1 (Strongly disagree) to 7 (Strongly agree). For each of the Big Five there were two items, of which one was reverse coded. After reversing these items, the measure of each of the 5 traits was the average of these two items.

Extraversion: 1, 6R

Agreeableness: 2R, 7

Conscientiousness; 3, 8R

Emotional Stability: 4R, 9

Openness to Experiences: 5, 10R

Short dark triad measure (Jones & Paulhus, 2014)

Please indicate how much you agree with each of the following statements.

Machiavellianism

1. It's not wise to tell your secrets.
2. I like to use clever manipulation to get my way.
3. Whatever it takes, you must get the important people on your side.
4. Avoid direct conflict with others because they may be useful in the future.
5. It's wise to keep track of information that you can use against people later.
6. You should wait for the right time to get back at people.
7. There are things you should hide from other people to preserve your reputation.
8. Make sure your plans benefit yourself, not others.
9. Most people can be manipulated.

Narcissism

1. People see me as a natural leader.
2. I hate being the center of attention. (R)
3. Many group activities tend to be dull without me.
4. I know that I am special because everyone keeps telling me so.
5. I like to get acquainted with important people.
6. I feel embarrassed if someone compliments me. (R)
7. I have been compared to famous people.
8. I am an average person. (R)
9. I insist on getting the respect I deserve.

Psychopathy

1. I like to get revenge on authorities.
2. I avoid dangerous situations. (R)
3. Payback needs to be quick and nasty.
4. People often say I'm out of control.
5. It's true that I can be mean to others.
6. People who mess with me always regret it.

7. I have never gotten into trouble with the law. (R)
8. I enjoy having sex with people I hardly know
9. I'll say anything to get what I want.

Coding

All items responded to with Likert scales from 1 (Strongly disagree) to 5 (Strongly agree). There are 27 items. Of these 27 items, 5 were reverse coded. After reversing these 5, the total score for each facet was the average of its 9 constituent items.

Receptiveness to opposing views (Minson et al., 2019)

The questions below address the manner in which you deal with contrary views and opinions on social and political issues that are important to you. When answering these questions think about the hotly contested issues in current social and political discourse (for example: universal healthcare, abortion, immigration reform, gay rights, gun control, environmental regulation, etc.). Consider especially the issues that you care about the most

1. I am willing to have conversations with individuals who hold strong views opposite to my own.
2. I like reading well thought-out information and arguments supporting viewpoints opposite to mine.
3. I find listening to opposing views informative.
4. I value interactions with people who hold strong views opposite to mine.
5. I am generally curious to find out why other people have different opinions than I do.
6. People who have opinions that are opposite to mine often have views which are too extreme to be taken seriously. (R)
7. People who have views that oppose mine rarely present compelling arguments. (R)

8. Information from people who have strong opinions that oppose mine is often designed to mislead less-informed listeners. (R)
9. Some points of view are too offensive to be equally represented in the media. (R)
10. Some issues are just not up for debate. (R)
11. Some ideas are simply too dangerous to be part of public discourse. (R)
12. I consider my views on some issues to be sacred. (R)
13. People who have views that oppose mine are often biased by what would be best for them and their group. (R)
14. People who have views that oppose mine often base their arguments on emotion rather than logic. (R)
15. Listening to people with views that strongly oppose mine tends to make me angry. (R)
16. I feel disgusted by some of the things that people with views that oppose mine say. (R)
17. I often feel frustrated when I listen to people with social and political views that oppose mine. (R)
18. I often get annoyed during discussions with people with views that are very different from mine. (R)

Coding

All items responded to with Likert scales from 1 (Strongly disagree) to 7 (Strongly agree). The 18 item scale contains 4 factors. The first contains items 15-18, the second items 1-5, and third items 6-8 and 13-14, and the fourth 9-12. We are primarily interested in the overall scale. Of these 18 items, 13 were reverse coded. After reversing these 11, the total score was the average of the 18 items.

The maximization inventory (Diab et al., 2008)

Please indicate how much each of the following statements sounds like you.

1. No matter what it takes, I always try to choose the best thing.
2. I don't like having to settle for "good enough".
3. I am a maximizer.
4. No matter what I do, I have the highest standards for myself.

5. I will wait for the best option, no matter how long it takes.
6. I never settle for second best.
7. I am uncomfortable making decisions before I know all of my options.
8. Whenever I'm faced with a choice, I try to imagine what all the other possibilities are, even ones that aren't present at the moment.
9. I never settle.

Coding

All items responded to with Likert scales from 1 (Not at all like me) to 5 (Very much like me). Total score an average of responses to the 9 items.

Attitude certainty (DeMarree et al., 2020)

Instructions: We are now going to ask you a series of questions about your attitudes towards a number of different issues/objects.

Using the scale below, please indicate your attitude towards

How certain are you of your attitude towards X?

1. Affirmative Action
2. George W. Bush
3. Paper Plates
4. Coffee
5. College Football
6. Mexican Food
7. Tide Laundry Detergent
8. Sunbathing (Tanning)
9. The Pope
10. Acid Rain

Coding

Respondents indicate their attitudes and certainty in those attitudes.

Attitudes: 9 point Likert from Extremely Negative to Extremely Positive

Certainty: 9 point Likert from Not at all Certain to Extremely Certain

The total score is the average of the 10 measures of attitude certainty.

Actively open-minded thinking (Baron, 2019)

This version is from the SJDM archive of measures. It may differ slightly from the

published version.

Please indicate how much you agree with each of the following statements.

1. True experts are willing to admit to themselves and others that they are uncertain or that they don't know the answer.
2. People should take into consideration evidence that goes against conclusions they favor.
3. Being undecided or unsure is the result of muddled thinking. (R)
4. People should revise their conclusions in response to relevant new information.
5. Changing your mind is a sign of weakness. (R)
6. People should search actively for reasons why they might be wrong.
7. It is OK to ignore evidence against your established beliefs. (R)
8. It is important to be loyal to your beliefs even when evidence is brought to bear against them. (R)
9. There is nothing wrong with being undecided about many issues.
10. When faced with a puzzling question, we should try to consider more than one possible answer before reaching a conclusion.
11. It is best to be confident in a conclusion even when we have good reasons to question it. (R)

<u>Coding</u>

All items responded to with Likert scales from 1 (Completely disagree) to 5 (Completely agree). Of the 11 items, 5 were reverse coded. After reversing these items, the total score was an average of responses to the 11 items.

Appendix B: Supplemental study

Preface

In Studies 1-2, we established that there are significant positive correlations among different elicitations of overconfidence, both in a single time period and across time. This supports the idea that there is a dispositional component to overconfident beliefs. However, our primary 12 different measures of overconfidence required participants to answer 204 questions (including both choices/estimates and confidence measures)—in this supplemental study we ask whether it is possible to measure core overconfidence by asking people just one measure of each elicitation of overconfidence (three measures total; 51 questions). We fixed these measures to each be from a different domain to sample overconfidence measures broadly as indicators of core overconfidence. This design is both more efficient (in terms of the number of questions) and a more conservative test of the correlations between different forms of overconfidence, given the noise added by the domains changing across individuals. However, the reduction in the number of overconfidence measures led to a great reduction in the reliability of our overconfidence measures. This impairs our ability to interpret the relationships observed in the study.

One key finding in Study 1 was that core overconfidence was strongly associated with individual differences—both in self-views and in cognitive styles. In this supplemental study, we probed several questions related to this finding. We first aimed

to replicate the associations we established in Study 1, in a larger, Nationally representative sample. Next, we included a wider range of related measures to test for other constructs that might be related to core overconfidence. We divided these measures into those associated with people's self-views and cognitive styles to establish whether one kind of individual difference was more strongly associated with core overconfidence. Yet the reduction in the reliability of our core overconfidence measure may have attenuated the nature of these associations with traits. As a result, we urge for caution in the interpretation of these data. The presence of a relationship suggests it is likely there is a strong relationship, but the absence cannot be seen as strong evidence for a null.

Methods

Participants

We recruited 1,157 participants from the survey platform ROI Rocket[18] in a pre-registered design (https://osf.io/bmgj5), with the sample targeting representing the U.S. population on age, gender, income, race, and ethnicity levels in line with the 2010 Census values.

Procedure

[18] Our original targeted sample size was N = 800, but the data at this stage was not close enough to being representative of the U.S. population on age, so we collected additional responses to balance out the age demographics. The final sample was within 10 percentage points of the 2010 Census values for each dimension except age, where older (55-64 and 65+) respondents were somewhat overrepresented and younger (18-34) participants underrepresented.

In the supplemental study, we moved away from asking all participants all of our overconfidence measures. We instead asked participants 3 measures of overconfidence out of the possible 12 domain-elicitation pairs with the aim of using each measure as a means to estimate core overconfidence. Each measure was a different elicitation and different domain. We chart the structure of the data in Table B.1 (Panel A). Each row of data refers to a participant—each of whom responded to a measure for each elicitation, which were fixed to be in different domains. There are 24 such combinations.[19] As a result, participants were randomly assigned to one of 24 conditions that determined the blocks of overconfidence measures they faced. These 3 blocks of items were presented in a randomized order.

[19] If you consider the elicitations as fixed, given that everyone answered one of each, you then have to choose 3 different domains from the total of 4, where order matters as it implies a different domain-elicitation pairing. Thus, the problem simply becomes 4P3 = 24.

Table B.1: The structure of the data in the supplemental study.

A	Overestimation		Overplacement		Overprecision	
Person	Domain	OE	Domain	OPL	Domain	
1	Geography	0.10	History	-0.10	Science	
2	History	0.20	Science	-0.05	Literature	
3	Science	0.05	Literature	-0.30	Geography	
…						
N	Literature	-0.05	Geography	-0.40	History	

B	Overestimation		Overplacement		Overprecision	
Person	Domain	OE	Domain	OPL	Domain	
1	*Geography*	*0.10*	History	-0.10	Science	
2	*History*	*0.20*	Science	-0.05	Literature	
3	*Science*	*0.05*	Literature	-0.30	Geography	
…						
N	*Literature*	*-0.05*	Geography	-0.40	History	

Note. Panel A of Table B.1 provides some example rows of how the overconfidence data is structured to demonstrate that each person answered an overestimation, overplacement, and overprecision measure, all of which were in different domains. This means that the correlation between a pair of measures (e.g., overestimation and overplacement) contains noise associated with domains changing between each row in different ways. Panel B demonstrates an example of the nature of our holdout sample testing: ignoring the data from overestimation, estimating core overconfidence using just overplacement and overprecision, and then correlating that core overconfidence with the overestimation measurements.

The constituent measures of the supplemental study were identical to those of Study 2 (with the exception that they did not include the alternate measure of overestimation). The basic idea of the supplemental study's design is that in their three measures, people provide information about their propensity to be overconfident across each elicitation and in three of the four domains. Information from other respondents

(i.e., the average overconfidence in each domain-elicitation pair) can be used to interpret this information, and recover a measure of core overconfidence without all participants having to answer the full set of questions. If this method is effective it offers the benefit of efficiency, making our measurement methods more accessible, and thus more broadly applicable to different research designs.

After responding to the overconfidence measures, participants answered a series of scales and measures. As in Study 1, all participants were asked the items from the narcissistic personality inventory (Ames et al., 2006; $\alpha = 0.75$), intellectual humility scale (Leary et al., 2017; $\alpha = 0.81$), need for closure decisiveness facet (Kruglanski et al., 1993; $\alpha = 0.79$), and the dominance facet of extraversion (Wiggins et al., 1988; $\alpha = 0.90$). In addition, all participants answered the 3 items from the Cognitive Reflection Test (CRT; Frederick, 2005) and 4 items from the Berlin Numeracy Test (BNT; Cokely et al., 2012).

As well as asking people these measures, participants were randomly assigned to answer 3/10 of the following 10 measures; need for closure (closed mindedness; Kruglanski et al., 1993; $\alpha = 0.65$), optimism (Scheier et al., 1994; $\alpha = 0.84$), self-efficacy (Chen et al., 2001; $\alpha = 0.94$), desirability of control (Burger & Cooper, 1979; $\alpha = 0.83$), a short Big Five measure (Gosling, Rentfrow, & Swann, 2003; α between 0.21 and 0.46), a short dark triad measure (Jones & Paulhus, 2014; α between 0.79 and 0.82), the receptiveness to opposing views scale (Minson et al., 2019; $\alpha = 0.86$), a maximizing scale

(Diab et al., 2008; $\alpha = 0.80$), a measure of attitude certainty (DeMarree et al., 2020; $\alpha = 0.92$), and a measure of actively open-minded thinking (Baron, 2019; $\alpha = 0.77$).

Finally, as an exploratory analysis, we also included some short items that aimed to capture overconfidence as efficiently as possible, such as, *"Suppose you're in a large group of people playing a new sport, Zoob Ball. You don't know how your ability compares to other people's. You play one game against a randomly selected person and win. What percentage of people do you think you're better than? (0-100%)".*

The overconfidence measures were coded the same as in Studies 1-2.

Results

We collected 1,157 sets of data. For overestimation and overplacement, all sets of data were complete. For overprecision, some people did not return any valid intervals (i.e., *all* of their intervals deviated from the requirements). These participants' measures for the overprecision measure were made NA (n = 9). The analyses we present are among the responses which were non-NA.

Nature of overconfidence

We first tabulated the average overconfidence in each elicitation and domain for the supplemental study, with Study 1's data included for comparison (Table B.2). We again observed mixed results for overestimation, consistent underplacement, and significant overprecision. The correlation between the average overconfidence across

domains and elicitations between the studies was $r = .92$ ($p < .001$). In short, similar results were observed.

Table B.2: Overconfidence in each domain and elicitation across Study 1 and the supplemental study.

		Study 1		Supplemental Study	
		N	Estimate	N	Estimate
Overestimation	Geography	200	8.5	278	5.4
	History	200	-0.1	273	-1.7
	Literature	200	5.0	319	0.1
	Science	200	8.8	287	-2.4
Overplacement	Geography	200	-0.9	285	-8.7
	History	200	-4.9	282	-2.4
	Literature	200	-17.0	281	-8.7
	Science	200	-16.4	309	-3.1
Overprecision	Geography	200	41.9	293	49.2
	History	200	33.4	307	33.1
	Literature	200	29.4	291	50.1
	Science	200	37.1	266	53.6

Next, we tested the correlations between overconfidence in each of the three elicitations in the supplemental study. Unlike in Study 1 and 2, participants only answered one measure of each, and all three of the measures were in different domains. To further complicate matters, the domains changed across rows. Consider Table B.1—when looking at overestimation and overplacement, between Person 1 and Person 2, the former's domain changed from geography to history, and the latter from history to science. When asking if people who generally display more overestimation also display more overplacement (i.e., the correlation), this factor adds noise to this relationship. To

partially account for this, as per our pre-registration, we standardized each measure within domain-elicitation before correlating the three measures for each participant. We still anticipated depressed magnitudes of the correlations due to the changing domains.

Despite the changing domains, we still observed significant positive correlations among the three measures; between overestimation and overplacement the result was strongest ($r = 0.26$, $p < .001$), whereas it was somewhat weaker but statistically significant for overestimation and overprecision ($r = 0.08$, $p = .009$) and among overplacement and overprecision ($r = 0.12$, $p < .001$).

If we look at the pairwise correlations when fixing the domain (e.g., the correlation between literature overestimation and science overplacement), there are some further observations. The correlations varied from -.21 (science overestimation and geography overprecision) to .45 (science overestimation and history overplacement), with 31/36 being positive. The cell sizes for these tests varied from 79 to 119. Overall, this study provides further evidence for an association of the different forms of overconfidence, finding significant positive correlations despite a shorter form design and varying domains across rows.

In running these analyses, it came to our attention that the correlations among different forms of overconfidence may vary as a function of demographic factors. Notably, it appears that overconfidence is less correlated across forms for older respondents. Whereas the correlations between overestimation and overplacement,

overestimation and overplacement, and overplacement and overprecision were .30, .24, and .14 for the age group 18-34, these correlations decreased to .12, .04, and .10 for respondents over the age of 65 (both cells were well-powered). This highlights the importance of attaining representative samples, as features of the population of study may interact with the relationships among forms of overconfidence. There was suggestive evidence that greater education might increase the degree of correlation among forms of overconfidence.

Measuring and validating core overconfidence

We next estimated core overconfidence using the same procedure as Study 1. In brief, we standardized the three overconfidence measures within domain-elicitation, pivoted them into long form, and entered them into a linear mixed effects model that only contained fixed and random intercepts. We measured *core overconfidence* by extracting the random intercepts from the model.

This measure correlated significantly with each of the three constituent measures; overestimation ($r = 0.68$, $p < .001$), overplacement ($r = 0.70$, $p < .001$), and overprecision ($r = 0.61$, $p < .001$). This within-sample correlation could be explained by noise: if each measure (e.g., overplacement) contains *actual* overplacement and noise, this noise will enter into the core overconfidence measure, which could create a correlation between core overconfidence and the overplacement measure. As a result, a stricter test is required. If we estimate core overconfidence with a subset of the data and

use it to predict an out-of-sample measure of overconfidence, the noise will be uncorrelated. We explore this possibility next.

To validate our measurement, we estimated core overconfidence with 2/3 of the elicitations and predicted people's responses in the third, holdout measure (see Panel B of Table B.1). We found that core overconfidence correlated with overestimation at .23 when dropping it from the estimation ($p < .001$), at .26 with overplacement when it was dropped ($p < .001$), and at .13 with overprecision when it was dropped ($p < .001$). In short, we were effectively able to drop each of the three elicitations, estimate core overconfidence, and predict into the dropped elicitation.

We next consider the metric we have been using to assess the reliability of our measurement of core overconfidence. Whereas the value of 1 – w (1 – the pooling coefficient) was around 0.85 in Studies 1-2, in the supplemental study it had a median value of 0.359. This indicates *greatly* reduced reliability in our measurement of overconfidence. This reduced reliability makes it difficult to interpret the strength of core overconfidence's relationships with other constructs.

Overconfidence and individual differences

In Study 1, we found strong relationships between narcissism, dominance, and intellectual humility with core overconfidence, also finding some evidence that need for closure (particularly the decisiveness facet) might predict overconfidence. In the supplemental study, we sought to replicate and build on these results. All respondents

were asked six individual difference measures, including the four from Study 1 listed above and two new ones; the Cognitive Reflection Test (CRT, Frederick, 2005) and the Berlin Numeracy Test (Cokely et al., 2012). We tested the correlations between our core overconfidence measure (the random intercepts from the mixed effects model) and each of these individual difference measures.

We replicated the significant positive correlations of narcissism ($r = 0.21$, $p < .001$) and dominance ($r = 0.15$, $p < .001$) with overconfidence, though the magnitudes of the correlations were smaller than in Study 1. We also replicated the significant negative correlation between decisiveness and overconfidence ($r = -0.06$, $p = .031$). The correlation with intellectual humility did not attain significance ($r = -0.01$, $p = .694$, vs. -0.35 in Study 1). Of the two new individual difference measures, both cognitive reflection ($r = -0.19$, $p < .001$) and numeracy ($r = -0.11$, $p < .001$) were significant negative predictors of overconfidence. Overall, we found more evidence that individual differences were associated with core overconfidence. Notably, narcissism, dominance, and decisiveness emerged as significant predictors across both studies, and CRT and BNT showed new promise.

In addition to these six measures that everyone answered, participants answered a random selection of three additional measures from a possible ten. The average cell sizes for these correlations was around 350. Some of the measures contained multiple

constructs, for a total of 16 additional individual difference measures. We discuss these next.

In addition to narcissism and dominance, we asked several additional self-view based measures. Notably, self-efficacy was significantly related with core overconfidence ($r = 0.19$, $p < .001$), as were all three elements of the Dark Triad (Jones & Paulhus, 2014). Machiavellianism ($r = 0.20$, $p < .001$), narcissism ($r = 0.26$, $p < .001$), and psychopathy ($r = 0.30$, $p < .001$) all displayed positive associations with core overconfidence. None of the Big Five were associated with core overconfidence; openness ($r = -0.04$, $p = .483$), conscientiousness ($r = 0.02$, $p = .707$), extraversion ($r = 0.10$, $p = .075$), agreeableness ($r = -0.07$, $p = .210$), and emotional stability ($r = -0.05$, $p = .310$) all failed to display relationships. Extraversion was relatively close, as might be expected from existing literature. Optimism ($r = -0.03$, $p = .584$), desire for control ($r = 0.08$, $p = .121$), and receptiveness ($r = -0.07$, $p = .179$) all did not display significant relationships with core overconfidence, though the latter two were in the predicted directions.

Moving to the measures of cognitive style; actively open-minded thinking negatively predicted core overconfidence ($r = -0.26$, $p < .001$), yet maximization had a positive relationship ($r = 0.12$, $p = .029$). We might have expected that those who engaged in more maximization would engage in more exhaustive information search, and thus display less overconfidence, but this surprising effect is aligned with the negative association between decisiveness observed across both studies—apparently making

decisions *quickly* doesn't necessarily mean people make them with overconfidence. Attitude certainty had a negative but insignificant correlation (r = -0.11, p = .058), and closed-mindedness showed a null result (r = 0.03, p = .536).

As a demonstration of the value of measuring *core overconfidence* rather than a specific instantiation, we include the signs and significances of the correlations between the individual difference measures and each of the three overconfidence elicitations in Table B.3. Choosing to elicit overprecision rather than overestimation would lead to different conclusions regarding the association between overconfidence and individual differences such as dominance. Another insight that comes from the table is that self-view based variables appear to be more important in determining overestimation and overplacement, whereas cognitive style based measures appear more important in determining overprecision.

Table B.3: Individual differences and correlations with specific types of overconfidence.

Type	Individual difference	Overestimation	Overplacement	Overprecision	Core
Self	Narcissism	+***	+***	+.	+***
Self	Dominance	+***	+***	-	+***
Cognitive	Decisiveness	-*	-	-	-*
Cognitive	Intellectual Humility	+	+	-**	-
Cognitive	CRT	-**	-.	-***	-***
Cognitive	Numeracy	-*	-	-***	-***
Self	Self-efficacy	+***	+.	+	+***
Self	Machiavellianism (DT)	+***	+**	+	+***
Self	Narcissism (DT)	+***	+***	+	+***
Self	Psychopathy (DT)	+***	+***	+	+***
Self	Openness	+	-	-	-
Self	Conscientiousness	-	+	+	+
Self	Extraversion	+*	+	+	+.
Self	Agreeableness	-.	-	-	-
Self	Emotional stability	-	-	-	-
Self	Optimism	+	-	-	-
Self	Desire for Control	+*	+.	-	+
Self	Receptiveness	-	-	-	-
Cognitive	AOT	-*	-***	-***	-***
Cognitive	Maximization	+***	+	-	+*
Cognitive	Attitude certainty	+	-	-***	-.
Cognitive	Closed mindedness	-	+	+.	+

Note. *** p < .001; ** p <.01; * p < .05; . p < .10

The individual difference measures that displayed significant relationships with core overconfidence fall into two main buckets: self-view variables that primarily inflated overconfidence through overestimation and overplacement (narcissism and the dark triad, dominance, and self-efficacy), and cognitive style variables that depressed

overconfidence, especially through overprecision (CRT, BNT, and AOT). It is possible that there are two factors within core overconfidence with one set of traits that is generally associated with overestimation and overplacement, and another cluster with overprecision. This could also help to explain the different findings regarding the contributions of confidence and accuracy to overconfidence in overprecision relative to the other forms of overconfidence. One relevant point is that exhibiting less overprecision may require some statistical sophistication, so the negative effect of cognitive style variables on that elicitation could partially reflect differences in respondents' *understanding* of the elicitation's task. Yet the strongest determinants predicted overconfidence in the same direction across all three elicitations (e.g., narcissism, CRT, AOT).

In sum, we found evidence that a wider set of individual differences are associated with overconfidence—both core overconfidence and its constituent measures. Among these relationships, both self-view and cognitive style measures emerged as key predictors, and appear to be important in influencing core overconfidence. It is not clear that one type of measure is *more* important in determining core overconfidence.

We next intended to investigate the possibility of predicting core overconfidence *even more efficiently* by introducing new items for this purpose, but this task was hampered by the reduced reliability of the core overconfidence measure. In short, several of these items displayed promise in predicting core overconfidence. Most

notably, the two responses to oci3 were strongly positively correlated with core overconfidence, as was the estimated performance percentile from oci5. In addition, oci1, oci4_1, oci4_4, and oci7_up showed positive relationships, and oci4_1, oci4_2, and oci7_depends showed negative relationships. The constituent measures of oci4 and oci7 are harder to interpret—they are negatively correlated within-measure by construction and thus it is difficult to discern exactly which choices are associated with overconfidence. Overall, the correlations ranged from -.11 to 0.23. We do not expand further on this as, given the reduced reliability of our measure of core overconfidence, the strength of the associations is difficult to interpret.

Discussion

The results of the supplemental study corroborate and extend those of Study 1 and 2—there is evidence for a core overconfidence that underpins each of the three forms, and is observed in a new study population. This is shown by the correlations among the different elicitations, and the ability of core overconfidence to predict out-of-sample measures of overconfidence. Further, core overconfidence is strongly associated with a variety of individual difference measures, both in terms of self-views and people's cognitive styles.

We found that our experimental design—where people saw three domain-elicitations—could efficiently access core overconfidence but that its measurement reliability was greatly reduced. This means that our correlations between core

overconfidence and individual difference measures are likely to have captured only the strongest relationships—thus, a significant effect should indicate a real relationship, but a non-significant one could be a false negative derived from this reduced reliability. The reduced reliability should not have affected the *order* of the magnitudes of relationships, but rather depressed them all. As a result, the rank order of associations observed for different variables may give an indication of how the strength of relationships would be ranked with more reliable measurement. Further work is needed to establish the full set of association of core overconfidence in a sample with more reliable measurement of this construct, such as in Studies 1-2.

References

Ackerman, R. A., Hands, A. J., Donnellan, M. B., Hopwood, C. J., & Witt, E. A. (2017). Experts' Views Regarding the Conceptualization of Narcissism. Journal of Personality Disorders, 31(3), 346–361. http://dx.doi.org/101521pedi201630254

Ahmed, A. S., & Duellman, S. (2013). Managerial Overconfidence and Accounting Conservatism. Journal of Accounting Research, 51(1), 1–30. https://doi.org/10.1111/j.1475-679X.2012.00467.x

Ames, D. R., & Kammrath, L. K. (2004). Mind-Reading and Metacognition: Narcissism, not Actual Competence, Predicts Self-Estimated Ability. Journal of Nonverbal Behavior, 28(3), 187–209. https://doi.org/10.1023/B:JONB.0000039649.20015.0e

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. Journal of Research in Personality, 40(4), 440–450. https://doi.org/10.1016/j.jrp.2005.03.002

Baron, J. (2019). Actively open-minded thinking in politics. Cognition, 188, 8–18. https://doi.org/10.1016/j.cognition.2018.10.004

Ben-David, I., Graham, J. R., & Harvey, C. R. (2007). Managerial Overconfidence and Corporate Policies (Working Paper No. 13711; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w13711

Ben-David, I., Graham, J. R., & Harvey, C. R. (2013). Managerial Miscalibration*. The Quarterly Journal of Economics, 128(4), 1547–1584. https://doi.org/10.1093/qje/qjt023

Blais, A.-R., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence judgments in comparative judgment tasks: The role of cognitive styles. Personality and Individual Differences, 38(7), 1701–1713. https://doi.org/10.1016/j.paid.2004.11.004

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. Organizational Behavior and Human Decision Processes, 49(2), 188–207. https://doi.org/10.1016/0749-5978(91)90048-X

Brown, R., & Sarma, N. (2007). CEO overconfidence, CEO dominance and corporate acquisitions. Journal of Economics and Business, 59(5), 358–379. https://doi.org/10.1016/j.jeconbus.2007.04.002

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. Journal of Personality and Social Psychology, 92(5), 938–956. https://doi.org/10.1037/0022-3514.92.5.938

Budescu, D. V., & Du, N. (2007). Coherence and Consistency of Investors' Probability Judgments. Management Science, 53(11), 1731–1744. https://doi.org/10.1287/mnsc.1070.0727

Budescu, D. V., Wallsten, T. S., & Au, W. T. (1997). On the Importance of Random Error in the Study of Probability Judgment. Part II: Applying the Stochastic Judgment Model to Detect Systematic Trends. Journal of Behavioral Decision Making, 10(3), 173–188. https://doi.org/10.1002/(SICI)1099-0771(199709)10:3<173::AID-BDM261>3.0.CO;2-6

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. Motivation and Emotion, 3(4), 381–393. https://doi.org/10.1007/BF00994052

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

Cain, D. M., Moore, D. A., & Haran, U. (2015). Making sense of overconfidence in market entry. Strategic Management Journal, 36(1), 1–18. https://doi.org/10.1002/smj.2196

Camerer, C., & Lovallo, D. (1999). Overconfidence and Excess Entry: An Experimental Approach. American Economic Review, 89(1), 306–318. https://doi.org/10.1257/aer.89.1.306

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

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. Organizational Research Methods, 4(1), 62–83. https://doi.org/10.1177/109442810141004

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. Judgement and Decision Making, 7(1), 25–47.

Cutler, B. L., & Wolfe, R. N. (1989). Self-monitoring and the association between confidence and accuracy. Journal of Research in Personality, 23(4), 410–420. https://doi.org/10.1016/0092-6566(89)90011-1

DeMarree, K. G., Petty, R. E., Briñol, P., & Xia, J. (2020). Documenting individual differences in the propensity to hold attitudes with certainty. Journal of Personality and Social Psychology, 119(6), 1239–1265. https://doi.org/10.1037/pspa0000241

Diab, D. L., Gillespie, M. A., Consulting, D., & Highhouse, S. (2008). Are maximizers really unhappy? The measurement of maximizing tendency. Judgment and Decision Making, 3(5), 7.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. Psychological Review, 101(3), 519–527. https://doi.org/10.1037/0033-295X.101.3.519

Ferrell, W. R. (1994). Discrete subjective probabilities and decision analysis: Elicitation, calibration and combination. In Subjective probability (pp. 411–451). John Wiley & Sons. https://doi.org/10.1038/scientificamerican1157-128

Frederick, S. (2005). Cognitive Reflection and Decision Making. Journal of Economic Perspectives, 19(4), 25–42. https://doi.org/10.1257/089533005775196732

Galasso, A., & Simcoe, T. S. (2011). CEO Overconfidence and Innovation. Management Science, 57(8), 1469–1484. https://doi.org/10.1287/mnsc.1110.1374

Gelman, A., & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review, 98(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

Glaser, M., Langer, T., & Weber, M. (2013). True Overconfidence in Interval Estimates: Evidence Based on a New Measure of Miscalibration. Journal of Behavioral Decision Making, 26(5), 405–417. https://doi.org/10.1002/bdm.1773

Goodman-Delahunty, J., Granhag, P., Hartwig, M., & Loftus, E. (2010). Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes. Psychology Public Policy and Law, 16. https://doi.org/10.1037/a0019060

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 24(3), 411–435. https://doi.org/10.1016/0010-0285(92)90013-R

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(4), 719–731. https://doi.org/10.1037/0278-7393.11.1-4.719

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. Journal of Personality and Social Psychology, 66(1), 206–219. https://doi.org/10.1037/0022-3514.66.1.206

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. Assessment, 21(1), 28–41. https://doi.org/10.1177/1073191113514105

Juslin, P. (1994). The Overconfidence Phenomenon as a Consequence of Informal Experimenter-Guided Selection of Almanac Items. Organizational Behavior and Human Decision Processes, 57(2), 226–246. https://doi.org/10.1006/obhd.1994.1013

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(4), 1038–1052. https://doi.org/10.1037/0278-7393.25.4.1038

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80(4), 237–251. https://doi.org/10.1037/h0034747

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. Psychological Review, 103(3), 582–591. https://doi.org/10.1037/0033-295X.103.3.582

Kang, S., & Kim, J. W. (2022). The Fragility of Experts: A Moderated-Mediation Model of Expertise, Expert Identity Threat, and Overprecision. Academy of Management Journal, 65(2), 577–605. https://doi.org/10.5465/amj.2019.0899

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It Depends on How, What, and Whom You Ask. Organizational Behavior and Human Decision Processes, 79(3), 216–247. https://doi.org/10.1006/obhd.1999.2847

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. Applied Cognitive Psychology, 15(3), 321–341. https://doi.org/10.1002/acp.705

Konrath, S., Meier, B. P., & Bushman, B. J. (2014). Development and Validation of the Single Item Narcissism Scale (SINS). PLOS ONE, 9(8), e103469. https://doi.org/10.1371/journal.pone.0103469

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 6(2), 107–118. https://doi.org/10.1037/0278-7393.6.2.107

Krueger, J. I., Freestone, D., & MacInnis, M. L. M. (2013). Comparisons in research and reasoning: Toward an integrative theory of social induction. New Ideas in Psychology, 31(2), 73–86. https://doi.org/10.1016/j.newideapsych.2012.11.002

Krueger, J. I., & Heck, P. R. (2020). The End of Overconfidence: Review of 'Perfectly confident: How to calibrate your decisions wisely' by Don A. Moor. https://doi.org/10.31234/osf.io/twrh5

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. Journal of Personality and Social Psychology, 77(2), 221–232. https://doi.org/10.1037/0022-3514.77.2.221

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. Journal of Personality and Social Psychology, 65(5), 861–876. https://doi.org/10.1037//0022-3514.65.5.861

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). Organizational Behavior and Human Decision Processes, 102(1), 76–94. https://doi.org/10.1016/j.obhdp.2006.10.002

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

Macenczak, L. A., Campbell, S., Henley, A. B., & Campbell, W. K. (2016). Direct and interactive effects of narcissism and power on overconfidence. Personality and Individual Differences, 91, 113–122. https://doi.org/10.1016/j.paid.2015.11.053

Malmendier, U., & Tate, G. (2004). Who Makes Acquisitions? A Test of the Overconfidence Hypothesis. NBER Working Paper, 73.

Malmendier, U., & Tate, G. (2005a). CEO Overconfidence and Corporate Investment. The Journal of Finance, 60(6), 2661–2700. https://doi.org/10.1111/j.1540-6261.2005.00813.x

Malmendier, U., & Tate, G. (2005b). Does Overconfidence Affect Corporate Investment? CEO Overconfidence Measures Revisited. European Financial Management, 11(5), 649–659. https://doi.org/10.1111/j.1354-7798.2005.00302.x

Malmendier, U., Tate, G., & Yan, J. (2011). Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies. The Journal of Finance, 66(5), 1687–1733. https://doi.org/10.1111/j.1540-6261.2011.01685.x

Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in Narcissism. Annual Review of Clinical Psychology, 13(1), 291–315. https://doi.org/10.1146/annurev-clinpsy-032816-045244

Minson, J. A., Chen, F. S., & Tinsley, C. H. (2020). Why Won't You Listen to Me? Measuring Receptiveness to Opposing Views. Management Science, 66(7), 3069–3094. https://doi.org/10.1287/mnsc.2019.3362

Moore, D. A. (2020). Perfectly confident: How to calibrate your decisions wisely. HarperCollins.

Moore, D. A., & Dev, A. S. (2018). Individual Differences in Overconfidence. In Encyclopedia of Personality and Individual Differences. http://doi.apa.org/getdoi.cfm?doi=10.1037/pspa0000018

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

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

Moore, D. A., & Swift, S. A. (2011). The three faces of overconfidence in organizations. In Social psychology and organizations (pp. 147–184). Routledge/Taylor & Francis Group. https://doi.org/10.4135/9781446250556.n9

Muthukrishna, M., Henrich, J., Toyokawa, W., Hamamura, T., Kameda, T., & Heine, S. J. (2018). Overconfidence is universal? Elicitation of Genuine Overconfidence (EGO) procedure reveals systematic differences across domain, task knowledge, and incentives in four populations. PLOS ONE, 13(8), e0202288. https://doi.org/10.1371/journal.pone.0202288

Oakley, J. (2021). SHELF: Tools to Support the Sheffield Elicitation Framework (1.8.0) [Computer software]. https://CRAN.R-project.org/package=SHELF

Olivares, O. J. (1994). Overconfidence and personality: Do individual differences exist in the way people make confidence judgments? [Ph.D., Texas A&M University]. https://www.proquest.com/docview/304066470/abstract/3C1E7D190D754237PQ/1

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

Oskamp, S. (1965). Overconfidence in case-study judgments. Journal of Consulting Psychology, 29(3), 261–265. https://doi.org/10.1037/h0022125

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The Role of Individual Differences in the Accuracy of Confidence Judgments. The Journal of General Psychology, 129(3), 257–299. https://doi.org/10.1080/00221300209602099

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. Journal of Personality and Social Psychology, 84(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Ramanaiah, N. V., & Goldberg, L. R. (1977). Stylistic Components of Human Judgment: The Generality of Individual Differences. Applied Psychological Measurement, 1(1), 23–39. https://doi.org/10.1177/014662167700100106

Russo, J., & Schoemaker, P. (1992). Managing Overconfidence. Sloan Management Review, 33, 7–17.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. Journal of Research in Personality, 38(5), 473–480. https://doi.org/10.1016/j.jrp.2003.09.010

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. Journal of Personality and Social Psychology, 67(6), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Soll, J. B. (1996). Determinants of Overconfidence and Miscalibration: The Roles of Random Error and Ecological Structure. Organizational Behavior and Human Decision Processes, 65(2), 117–137. https://doi.org/10.1006/obhd.1996.0011

Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(3), 780–805. https://doi.org/10.1037/a0015145

Soll, J. B., Palley, A. B., & Rader, C. A. (2021). The Bad Thing About Good Advice: Understanding When and How Advice Exacerbates Overconfidence. Management Science. https://doi.org/10.1287/mnsc.2021.3987

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

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. Journal of Experimental Psychology: General, 127(2), 161–188. https://doi.org/10.1037/0096-3445.127.2.161

Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment Extremity and Accuracy Under Epistemic vs. Aleatory Uncertainty. Management Science, 63(2), 497–518. https://doi.org/10.1287/mnsc.2015.2344

Teigen, K. H., & JØrgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. Applied Cognitive Psychology, 19(4), 455–475. https://doi.org/10.1002/acp.1085

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

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

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R). Multivariate Behavioral Research, 23(4), 517–530. https://doi.org/10.1207/s15327906mbr2304_8

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective Probability Intervals: How to Reduce Overconfidence by Interval Evaluation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(6), 1167–1175. https://doi.org/10.1037/0278-7393.30.6.1167

Wolfe, R. N., & Grosch, J. W. (1990). Personality Correlates of Confidence in One's Decisions. Journal of Personality, 58(3), 515–534. https://doi.org/10.1111/j.1467-6494.1990.tb00241.x

Wright, A. G. C. (2016). On the Measure and Mismeasure of Narcissism: A Response to "Measures of Narcissism and Their Relations to DSM-5 Pathological Traits: A Critical Reappraisal." Assessment, 23(1), 10–17. https://doi.org/10.1177/1073191115599054

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. Journal of Experimental Psychology: General, 124(4), 424–432. https://doi.org/10.1037/0096-3445.124.4.424

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. Journal of Behavioral Decision Making, 10(1), 21–32. https://doi.org/10.1002/(SICI)1099-0771(199703)10:1<21::AID-BDM243>3.0.CO;2-G

Biography

Matthew Asher Lawson received a bachelor's degree in Economics and Management from Merton College of the University of Oxford (2018), graduating with first-class honors. He has published several journal articles; "Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules", "Of Pandemics, Politics, and Personality: The Role of Conscientiousness and Political Ideology in Sharing of Fake News", "Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language", as well as an encyclopedia entry "Judgment and Decision-Making Processes" and a book chapter, "Saying More Than We Know: How Language Provides a Window into the Human Psyche". He received a National Science Foundation "Innovations in Graduate Education" grant (2019), as well as winning the Best Student Paper Award at the International Association for Conflict Management and being a finalist in the Decision Analysis Society's student paper award.