Fall 11-14-2023

Do I Listen to You, or Do I Listen to Me? An Individual Difference Investigation into Advice Utilization

Danielle Nicole Sanchez-Combs
University of New Mexico - Main Campus

Danielle Sanchez-Combs
Candidate

Psychology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Dr. Eric Ruthruff, Ph.D., Chairperson

Dr. Tania Reynolds, Ph.D., Dissertation Committee

Dr. Jeremy Hogeveen, Ph.D., Dissertation Committee

Dr. M. Lee Van Horn, Ph.D., Dissertation Committee

Do I Listen to You, or Do I Listen to Me?

An Individual Difference Investigation into Advice Utilization

by

Danielle Sanchez-Combs

B.S., Psychology, University of New Mexico, 2018

M.S., Psychology, University of New Mexico, 2021

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Psychology

The University of New Mexico
Albuquerque, New Mexico

December, 2023

Do I Listen to You, or Do I Listen to Me?

An Individual Difference Investigation into Advice Utilization

By

Danielle Sanchez-Combs

B.S., Psychology, University of New Mexico, 2018

M.S., Psychology, University of New Mexico, 2021

Ph.D., Psychology, University of New Mexico, 2023

ABSTRACT

This work addresses three fundamental questions. First, can the source of the advice (crowd or single advisor) be leveraged to enhance advice use? Second, does high skill and high metacognitive ability predict greater advice use or are these individuals also blind to the need for advice? Finally, can *personality*, *performance*, and *pre-advice confidence* factors be used to profile those most likely to benefit from advice? Results indicated surprisingly low advice taking rates (∼25% – ∼26%) from both advisors, despite the advice being 100% accurate. Advice taking was even lower when individuals were in a high-confidence state, with high-skilled individuals taking advice 18% of the time they made an error, compared to 7% for the low-skilled. Significant individual differences in advice use were found, with those high in normative social influence more likely to take advice.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Overview of Aims

You've just made a big mistake – but you have not noticed it yet. A helpful colleague points out that something seems off with your recent results and encourages you to try a different approach. Is your colleague's advice enough of a "nudge" to get you to go back and review your procedures, enabling you to catch your error? Or do you carry on with your initial course of action, remaining blissfully unaware of your error until it's too late?

Individuals often receive advice like this in their day-to-day lives. However, previous research has found relative resistance to accepting advice, with advice use rates as low as 20% to 30%, even when advisors demonstrate high accuracy levels (i.e., 75% to 80%; Harry & Fischer, 1997; Soll & Larrick, 1999; Soll & Larrick, 2009; Duan, Gu, & Sun, 2016; Pescetelli, Hauperich, & Yeung, 2021). This resistance to taking advice persists even in 'high-risk' situations, where resources or money are at stake or when advisors possess more experience than the advice seeker (Harry & Fischer, 1997; Soll & Larrick, 1999; Soll & Larrick, 2009; See et al., 2011; Pescetelli, Hauperich, & Yeung, 2021). In the real world, unwillingness to accept valuable feedback has led to failure to avoid financial losses (Lewis, 2018), failure to prevent malware attacks (Malkin et al., 2017), and failure to prepare for an impending tornado (Ripberger et al., 2015).

While ignoring advice is prevalent, it might not be a universal human characteristic. Previous studies have found a high degree of variation in individuals' willingness to accept advice, with some individuals never taking advice and others taking advice as much as 72% of the time (Pescetelli, Hauperich, & Yeung, 2021). The primary goal of the present work is to ascertain why some individuals rarely accept advice (even when they should) and why others are better at using advice to correct their mistakes.

State-based factors, categorized as elements that fluctuate depending on the situation or social context, have been empirically linked to advice taking. Many state-based factors, such as emotion, self-esteem, "state-level narcissism," and a personal sense of power, have been found to influence individuals' willingness to use advice (Bailey et al., 2021; See et al., 2011; Kausel et al., 2015; De Wit et al., 2017). However, the link between trait-based factors and advice use are not well understood, a fact which has been highlighted by several advice-taking researchers (e.g., Bonaccio & Dalal, 2006; Duan, Gu, & Sun, 2016; Bailey et al., 2022). Yet, personality traits have been consistent predictors of other aspects of occupational and academic performance (e.g., Barrick, Mount, & Judge, 2001; Zell & Lesick, 2022). Thus, it is essential to explore how personality traits relate to advice taking, as it may have real-world implications for high-stakes professions like finance, emergency response, and aviation.

In addition to personality traits, it is critical to examine individual differences in metacognitive ability, such as being able to detect when one is likely to be correct versus incorrect (i.e., metacognitive sensitivity; Fleming & Lau, 2014). Previous studies on advice taking have considered trial confidence as a predictor of advice use, a common component in metacognition investigations (see Fleming & Lau, 2014). However, these studies have not explored whether certain individuals are better at detecting their errors and using advice to correct them. Therefore, the present research incorporates techniques from the fields of metacognition (e.g., Fleming & Lau, 2014), overconfidence (e.g., Moore & Schatz, 2017), and the Dunning-Kruger effect (e.g., Kruger & Dunning,

1999) to examine whether some individuals can more effectively use advice to catch and correct errors.

The final area of exploration in this work is whether individual differences lie in advice taking based on preference for a specific type of advisor (e.g., single versus multiple advisors). In traditional advice taking paradigms, a single advisor provides advice (e.g., Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; Harvey et al., 2000; Meshi et al., 2012; Hütter & Ache, 2016; Schultze et al., 2017; Tzini & Jain, 2018; Haran & Shalvi, 2020; Schultze & Loschelder, 2021). Research in other fields, such as social conformity, has found that the number of individuals offering advice can affect its acceptance, even when the advice is incorrect (Farjam & Loxbo, 2023; Bond & Smith, 1996; Hertz & Wiese, 2018). In traditional advice taking paradigms, the social influence of a single advisor may be insufficient to elicit advice taking, as a group size of at least three is needed to generate a conformity response (Wijenayake et al., 2021). This work will explore advice taking when the advice comes from a single peer (an advisor) versus a group of peers (a crowd) using a within-subjects design to understand the role of advisor preference on advice use.

The benefit of this research is twofold. First, studying individual differences in advice taking can provide insights into the mechanisms underlying the unfortunate rejection of good advice, which is relevant to the use of advice recommendation systems such as those powered by Artificial Intelligence (AI: Sutherland, Harteveld, & Young, 2016). This knowledge could lead to practical implications for personnel selection and training in high-stakes professions that require integrating one's judgment with others'. Second, results of this study can inform advice delivery. If advice use is enhanced when coming from a group of individuals (i.e., a “crowd”), then advice recommendation systems can leverage social consensus to improve advice taking.

Summary of Aims

The primary aim of this work is to explore individual differences in advice taking. This work brings the novel contributions of exploring the role the type of advisor (single versus multiple), personality, metacognition, skill, and overconfidence play in leading individuals to use (or not use) advice. Should this investigation be successful, the field will be able to better identify those who are most willing to accept advice and to optimize the design of advice/recommender systems.

Chapter II: Review of the Advice Taking Literature

The goal of this chapter is to review current research on whether and when people take advice. Thus, this chapter includes an overview of current leading theories of advice taking and how advice taking is currently measured in the literature. This examination is followed by an overview of advisor and advice-taker related factors that are thought to enhance or degrade advice taking.

Understanding Advice Taking: The Judge-Advisor System

To study advice taking, researchers have typically used the judge-advisor system paradigm. In a judge-advisor paradigm, one participant (the judge) performs decision-making duties, while the other (advisor) gives advice. Typically, the advice comes from an advisor that is in the room with the judge and concurrently completes the study with the judge (e.g., Harvey & Harries, 2004; Harvey & Fischer, 1997; Bonaccio & Dalal, 2006, Soll & Larrick, 2009). However, alternations to this design have used banks of previous participants' answers or simulated judgments from a fictious advisor (e.g., Harvey, Harries, & Fischer, 2000; Yaniv & Milyavsky, 2007; Pescetelli, Hauperich, & Yeung, 2021).

A prototypical advice taking study, as outlined in a comprehensive review of advice taking by Bonaccio and Dalal (2006), has three phases. First, participants are asked to provide a numerical estimate, such as estimating an individual's weight from a photo or the date of a historical event, as well as an estimate of how accurate they believe their judgment is (i.e., a confidence interval or rating; e.g., Yaniv & Milyavsky, 2007; Pescetelli, Hauperich, & Yeung, 2021). Next, participants receive advice from the advisor. The advisor's characteristics are often manipulated in terms of gender, confidence, and/or level of training or expertise in the domain (e.g., Harvey, Harries, & Fischer, 2000; Harvey & Fischer, 1997; Bailey et al., 2021; Pescetelli, Hauperich, & Yeung, 2021).

After receiving advice, participants then provide a final numerical estimate and sometimes a post-decision confidence interval or rating (e.g., Soll & Larrick, 2009; Van Swol & Sniezek, 2005; Pescetelli, Hauperich, & Yeung, 2021; Price & Stone, 2004). The distance between an individual's final estimate and the advisor's recommendation usually serves as a key outcome measure (e.g., Harvey & Harries, 2004; Harvey & Fischer, 1997; Yaniv, 2004; Yaniv & Milyavsky, 2007; Yaniv & Kleinberger, 2000; Soll & Larrick, 2009; Pescetelli, Hauperich, & Yeung, 2021). These calculations will be described in-depth in the next section.

Measuring Advice Taking: Utilization, Discounting, & Weight on Advice

When individuals accept advice, this is known as *advice utilization* (Bonaccio & Dalal, 2006). On the flip side, *advice discounting* is the degree to which individuals do not adjust their judgments to incorporate advice (Bonaccio & Dalal, 2006). To determine the degree to which individuals utilize or discount advice, judge-advisor system paradigms often use "weight on advice" (WOA) as the dependent variable (Bonaccio & Dalal, 2006). WOA can be summarized by the following proportion:

$$WOA = \frac{|\ Final\ Estimate - Initial\ Estimate\ |}{|\ Advised\ Estimate - Initial\ Estimate\ |}$$

WOA values reflects how far the advice taker has moved from the initial estimate (WOA = 0) to the advisor's estimate (WOA = 1) (Bailey et al., 2022). A recent meta-analysis conducted by Bailey et al. (2022) found that, across 346 effect sizes from 53 studies, the average weight on advice was only .39, indicating a low weight on the advice of others.

Alternatively, "weight on own estimate" (WOE) is sometimes used over WOA. WOE can be expressed as:

$$WOE = \frac{|\, Advised\ Estimate - Final\ Estimate\, |}{|\, Advised\ Estimate - Initial\ Estimate\, |}$$

A WOE value of .5 indicates equal weight on the advice and one's initial estimate, while values closer to 1 indicate a higher weight on one's own advice and values closer to 0 indicate low weight on one's own advice (Yaniv & Kleinberger, 2000).

Why Do Individuals Utilize or Discount Advice?

The central question of advice-taking researchers has been why individuals do not accept good advice when it is offered to them. Current theories in the advice-taking literature highlight three suspected causes for why individuals discount advice. *Consistency theory* argues that individuals make decisions that are consistent with their internal attitudes, values, and personal histories (Russo et al., 1996; Simon & Holyoak, 2002). Specifically, when presented information conflicts with an individual's prior decisions, the resulting *cognitive dissonance* is resolved through *cognitive distortion*. Cognitive distortion is the attendance to information that matches one's initial opinion and the change or omittance of information that opposes it (Russo et al., 1996; Simon & Holyoak, 2002). Cognitive dissonance is defined as the discomfort experienced when holding two conflicting opinions (Russo et al., 1996). Prior experiments demonstrate that, in situations for which individuals are asked to choose between alternatives, such as two restaurants or consumer purchases (e.g., backpacks), individuals will favorably weigh attributes (e.g., location, color, size, etc.) of the choice they have selected over comparable alternatives (Russo et al., 1996; Russo et al., 1998). Therefore, the desire to maintain cognitive consistency acts as an *anchoring and adjustment* function for which individuals settle on a decision and have trouble incorporating new information. Thus, to eliminate cognitive dissonance provoked by the advisor's recommendation conflicting with theirs, individuals disregard the advice, regardless of how useful it is.

The second prevailing theory is based on the *self/other effect*, in which a decision-maker has access to the evidence they used to make their decision but not the evidence the advisor used (Yaniv, 2004). Due to this lack of clarity regarding others' judgments, individuals reject the advice in favor of their own judgment.

The final theory is the *egocentric bias* account, which stipulates that individuals prefer their judgment over others' judgments due to the belief that their decision-making capabilities are superior (Krueger, 2003; Bonaccio & Dalal, 2006). Cited evidence to support the egocentric bias account highlights that individuals typically place low weight on advice, adjusting their estimates by as little as 20% to 30% towards the advice giver's estimation (Bonaccio & Dalal, 2006; Soll & Larrick, 2009). However, confounding factors such as accuracy of the advice, participants' perception of the advisor's accuracy, and participants' perception of their expertise are often not accounted for (e.g., Study 1: Yaniv & Kleinberger, 2000; Yaniv, 2004; Yaniv & Milyavsky, 2007; Soll & Larrick, 2009). Thus, a true test of the egocentric bias account is still missing from the literature.

Factors Known to Influence Advice Taking

To understand when individuals are likely to take or ignore advice, it is important to consider two key factors: factors related to the advisor, such as their experience and the quality of their advice, and factors related to the judge, such as their age, gender, environment, and personality.

Factors of the Advisor

Advisor Expertise. One of the most explored factors known to influence whether individuals take advice is advisor expertise. Advisors can be described as experts, novices, or in neutral terms (i.e., a peer; Bailey et al., 2022). Expertise can be qualified either through education or job experience (e.g., medical professionals) or task specific training (e.g., receiving more training trials than the judge). Consistently, individuals have been found to be more willing to take advice from experts than from novices (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; Sniezek et al., 2004; Bonnaccio & Dalal, 2006; Yaniv & Milyavsky, 2007; Bailey et al., 2021; Bailey et al., 2022).

A recent meta-analysis found that when advice came from novices, individuals took advice 32% of the time on average (across 40 effect sizes; Bailey et al., 2022). When the advisor was considered to be neutral in terms of their expertise, individuals took advice 38% of the time on average (across 170 effect sizes; Bailey et al., 2022). However, when the advisor was considered an expert, advice was utilized on average 48% of the time (across 80 effect sizes; Bailey et al., 2022). Overall, it is important to note that regardless of the expertise of the advisor, advice utilization was low across the board, occurring only 32% to 48% of the time.

Advice Quality. Advice quality has been the subject of many empirical investigations designed to ascertain how advice taking can be increased. Advice quality is the degree to which the presented advice is seen as accurate (Gardner & Berry, 1995; Yaniv & Kleinberger, 2000; Pescetelli, Hauperich, & Yeung, 2021). It is important to note, however, that several advice taking investigations do not constrain the accuracy of the advice, but rather choose advice randomly from a pool of prior participants' answers. Therefore, weight on advice outcome measurements in several studies may be contaminated by participants' ability to detect poor quality advice (e.g., Study 1: Yaniv & Kleinberger, 2000; Yaniv, 2004; Yaniv & Milyavsky, 2007; Soll & Larrick, 2009; Kausel et al., 2015; Duan, Gu, & Sun, 2016; De Wit et al., 2017). Previous research has highlighted that participants, prior to receiving direct feedback regarding the advisor's performance, are able to detect which advisors are poor versus more accurate (e.g., Harvey, Harries, & Fisher, 2000). Therefore, if the advice is not of good quality (i.e., accurate), then this may account for the low weight on advice seen in many studies. Consequently, the current work will improve upon previous work by utilizing 100% correct advice to avoid confounds from advice quality.

An investigation conducted by Harvey, Harries, and Fischer (2000) examined participants' ability to detect the accuracy of advice they had been given. Advice was generated by selecting a series of sales forecasts and adding random noise to make the advisors more or less accurate. Participants were informed after they had provided their final answer what the correct estimate was, allowing them to ascertain an advisor's accuracy over time. Participants were much better at assessing the accuracy of the advisor than using their advice. Correlations between participants ratings of an advisor's accuracy (rel PI) and their true accuracy (rel UI_e) had high correspondence, $r = .77$, $p < .05$. However, participants' ability to rate the advisors' accuracy (rel PI) and participants' use of the advice (rel UI_j) had a lower correlation, $r = .45$, $p < .05$, indicating that individuals were less likely to use an advisors' advice, even when viewing it as accurate.

Similarly, Yaniv and Kleinberger (2000) investigated the role of accuracy in participants' willingness to take advice. Advice was generated from a pool of prior participants' answers, using either the best performing prior participant (accuracy rate: 93%) or the worst performing participant (accuracy rate: 13%). The experimental group were told the correct answer each trial (i.e., were given feedback), allowing them to assess their and the advisor's error rate. When the advisor provided good advice, mean WOE for the feedback condition was M = .42 (higher weight on advisor's estimate), while WOE in the no feedback condition was M = .54 (near equal weighting of one's initial answer and the advisor's answer). Mean WOE for the advisor who provided poor advice in the feedback condition was M = .72 while WOE for the no feedback condition WOE was M = .73, indicating participants placed higher weight on their own estimates. When comparing the mean WOEs in the no feedback condition with the good advisor (M = .54) and poor advisor (M = .73), it is evident that individuals were able to determine the quality of the advice without assistance.

However, Schultze, Mojzisch, and Schulz-Hardt (2017) claimed to counter the results of Harvey, Harries, and Fischer (2000) and Yaniv and Kleinberger (2000) regarding the assertion that participants will completely disregard poor advice. Schultze and colleagues theorize that, regardless of the accuracy of the advice, individuals become anchored to advice presented to them and thus move towards it, even when the advice is inaccurate. In this study, participants were shown advice from three advisors: a highly competent one, a moderately competent one, and an incompetent one. Participants who received advice from the incompetent advisor were told that the advisor's estimates were no better than chance. Results indicate that participants' advice taking was a WOA of .50 for the competent advisor (equal weighting between one's initial answer and the advice), .18 for the moderately competent advisor (high preference for one's initial answer), and .008 for the incompetent advisor (very high preference for one's initial answer). A WOA of .008, $t(27) = 3.30$, $p = .003$, $d = 0.62$, for the incompetent advisor indicates a slight use of the advisor's estimate even when the advisor is poor.

In addition, individuals have been found to use preciseness of advice as a proximal cue to determine an advisor's expertise. Schultze and Loschelder (2021) examined the role numerical precision plays in individuals' willingness to take advice, which was either induced by providing a precise estimate (e.g., Santa Fe is 70.25 miles away from Albuquerque) or a rounded value (e.g., Santa Fe is 70 miles away from Albuquerque), with precise estimates leading to greater advice utilization, $r = .17$, $p < .001$.

Finally, the distance of the advice from one's initial estimate has also been investigated as a factor in individuals' willingness to utilize advice. Hütter and Ache (2016) found that individuals are more willing to accept advice if the estimate is further away from their initial judgment, then if it is closer, with WOA increasing by $b = .12$ (i.e., a change in WOA from $M = .05$, $SD = .13$ to $M = .17$, $SD = .17$). However, despite this shift towards distant advice, participants still largely prefer their own advice (i.e., WOA is closer to 0 than to 1). However, in a multi-advisor paradigm for which the advisors did not reach a consensus, Yaniv and Milyavsky (2007) and Pescetelli and Yeung (2021) found that participants were more likely to choose advice from advisors whose estimate was closest to the participants' initial estimate.

Number of Advisors. In typical judge advisor paradigms, only one advisor presents advice to participants. Only one identified study to-date has investigated how advice taking is influenced by multiple advisors providing the same advice. In the first experiment, Mannes (2009) examined whether individuals believe advice from a group of two individuals is less accurate than advice from groups of nine individuals, finding that individuals believe groups of nine to be more accurate. However, his second experiment demonstrated that while participants believed larger groups to be more accurate, this did not predict their actual advice utilization.

In Mannes's (2009) second experiment, he examined whether individuals are more likely to adjust their judgments (i.e., temperature estimates) if the estimate comes from a group of one, two, four, or nine other individuals. He found that individuals were more likely to amend their temperature estimates when the group was comprised of two or four individuals, compared to groups of one or nine, as measured by improvement percentage in participants' mean absolute deviation (MAD) between their final estimates and the true temperature estimate. MAD scores were used as a dependent variable in the study to indicate the degree to which participants took the advice from the advisor and are lower in the single and nine advisor conditions. However, 67 of the 80 participants still favored their initial judgment over that of the group's, regardless of the group's size.

In contrast to Mannes's findings, prior research in the field of social conformity has consistently demonstrated that the higher the number of individuals providing a recommendation, the more likely individuals are to take the recommendation (Farjam & Loxbo, 2023; Bond & Smith, 1996; Asch, 1956; Klucharev et al., 2009; Perfumi et al., 2019; Wijenayake et al. 2020). I hypothesize that one reason why participants were more reluctant to take advice from the larger group in Mannes's study is due to uncontrolled advice quality. The accuracy of the groups' answers in Mannes's experiments was not controlled for. Therefore, rather than being a clear measure of whether or not individuals use advice when it comes from a larger group, it is confounded by individuals' ability to detect whether the advice is accurate or not. Poor advice quality in the larger group of nine individuals could account for the lower utilization rates. Therefore, the current work improves upon Mannes methodology in two ways. In the current work, participants will experience advice from both a single individual and a group of individuals who provide

100% accurate advice. The power of this is to explore whether some individuals favor one advisor type over the other, while controlling for differences in advisor accuracy.

In other advisor paradigms that have utilized more than one advisor (e.g., two to ten; Budescu & Rantilla, 2000; Yaniv & Milyavsky, 2007), the researchers deliberately ensured that the advisors signaled conflicting choices, precluding a clear consensus. Therefore, the increase in number of advisors was not found to significantly increase advice taking, with one study finding that decision-makers stuck with their initial decision 38% (two advisor condition), 39% (four advisor condition), and 40% (eight advisor condition) of the time (Yaniv & Milyavsky, 2007). Even on the 60% of trials in which individuals did not outright reject the advice, WOA rates indicated that participants were still heavily weighed their original answer (i.e., WOA rates of approximately .17). In addition, participants chose to utilize the advice of an advisor closest to their initial opinion (which they termed egocentric trimming; Yaniv & Milyavsky, 2007).

Social Conformity Research & the Number of Recommenders. Despite Mannes' finding that individuals use less advice when the group size is bigger, research in the field of social conformity has found that the larger the crowd of advisors, the more likely people are to adopt the crowd's answer (Farjam & Loxbo, 2023; Bond & Smith, 1996; Asch, 1956; Klucharev et al., 2009; Perfumi et al., 2019; Wijenayake et al. 2020). Higher conformity with a group's advice is thought to occur due to the impact of social norms, which are unspoken rules that govern an individual's behavior in a specific situation (Fehr & Fischbacher, 2004; Neville et al., 2021; Cialdini & Goldstein, 2004). Individuals are thought to conform to social norms to improve their accuracy and/or to be accepted by the group (Kuan, Zhong, & Chau, 2014; Hertz & Wiese, 2018; Klucharev et al., 2009;

Kastanakis & Balabanis, 2012; Neville et al., 2021; Klucharev et al., 2009; Cialdini & Goldstein, 2004). The likelihood of conformity increases when the crowd voicing their opinion is larger, the crowd is “in-group” or in an individual’s perceived social group, when the stimuli is ambiguous, or when the participant is female (Larsen, 1990; Bond & Smith, 1996; Salomons et al., 2021).

However, it is unclear whether this level of social influence will occur in advice-taking paradigms, which differ in important ways from conformity experiments. The advice-taking literature primarily asks individuals to decide prior to receiving advice, while the social conformity literature asks individuals to make decisions at the same time that they receive advice. Individuals are thought to “anchor” to an opinion when they are forced to decide prior to receiving advice. This anchoring occurs to maintain internal consistency in their attitudes and behaviors, leading to distortion of presented information that conflicts with their initial opinion (Russo et al., 1996; Simon & Holyoak, 2002). When advice is presented at the same time as the decision-maker is asked to decide, or beforehand, it is hypothesized to lead to individuals deciding based on heuristic judgment (e.g., recognition heuristic) rather than through analytical processes, which can lead to increased decision-making errors (Evans, 2003; De Neys & Glumicic, 2008; Gigerenzer & Goldstein, 2011).

Factors Related to the Judge

Factors related to the judge that have been hypothesized to affect a judge’s willingness to accept advice include their age and gender, emotions, self-esteem and personal sense of power, and personality factors. (Bailey et al., 2021; See et al., 2011; Kausel et al., 2015; De Wit et al., 2017). In the sections below, each of these individual differences will be explored in-depth.

Age and Gender. As part of a meta-analytic investigation, Bailey and colleagues (2022) hypothesized that the relationship between age and advice will follow a U-shaped curve: advice taking is theorized to be high in children and teens and those who are 65 or older, but low in individuals 18 – 64 years of age. However, only three studies were found to use participants under the age of 18 or over the age of 65. Therefore, the research team were unable to detect differences in advice utilization amongst these groups. Of the studies that examined participants 18 – 64 years of age, no age differences were detected.

Previous research has found differences in men and women's willingness to take advice. See (2011) found that women were overall less confident in their judgments and were more willing to take advice from their coworkers than men were. However, a meta-analysis conducted by Bailey (2022) failed to find any effect of participants' gender on propensity to take advice.

Emotions, Self-Esteem, and Narcissism. Work by de Hooge, Verlegh, and Tzioti (2014) examined the role of positive valenced emotions (e.g., love, gratitude, pride, or satisfaction) with negatively valenced emotions (e.g., anger, fear, shame, or guilt) that were either outwardly-focused (i.e., towards others) or internally-focused (i.e., towards self). Those who had experienced the positively valenced, outward-facing emotional manipulation (i.e., love or gratitude) were more likely to take advice (M = .59[1], SD = .23; Experiment 1) than those who received negatively valenced, outward-facing manipulation (i.e., anger or fear) (M = .29, SD = .23; Experiment 1). On the other hand, those who had positively valenced, inward-facing manipulation (i.e., pride or

[1] Advice taking was measured on a scale of 0 to 1, with 1 indicating that participants fully adjusted to the advisor's recommendation, while 0 means that fully rejected the advisor's recommendation.

satisfaction), were less likely to take advice (M = .44, SD = .33; Experiment 2) than those who were exposed to the negatively valenced, inward-facing manipulation (i.e., shame or guilt) (M = .63, SD = .24; Experiment 2).

A person's perceived sense of power (i.e., a positively valenced, inwardly-facing emotion such as pride) has been found to inhibit advice taking (See et al., 2011). Participants who have greater self-perceived power in an organization or have been primed to have an increased sense of personal power were found to be less willing to take advice from others, are less accurate in their estimates, and were more confident in their answers.

Similar results were found by Kausel et al. (2015). When a sense of power and confidence was primed (termed state-level narcissism in the paper), individuals were less likely to take advice, with state-level narcissism accounting for 33% of the variability in advice taking. Trait-level narcissism only accounted for 1% of the variability in advice taking. However, when individuals believe their personal power conveys a sense of personal responsibility to those around them, they are more likely to take advice from others than those who view their perceived power as an opportunity for personal gain (De Wit et al., 2017).

Finally, self-esteem, which is also related to concepts of personal power, has been found to affect individuals' willingness to accept advice. A study conducted by Duan, Gu, and Sun (2016) found that individuals with higher self-reported self-esteem were more likely to discount advice than their lower self-esteem peers.

Relatedly, clinical disorders affecting self-image, self-esteem, and perceived personal power have also been found to play a role in individuals' advice taking. Hofheinz and colleagues (2017) found that those who suffered from depression were *more* likely to take advice than their healthy peers. Those who did not suffer from depression made only a 20% – 30% shift towards the advisor's answer. However, those who were depressed shifted 40% towards the advisor's answer, a 10% higher shift than healthy controls. Similarly, those who were experiencing emotional ambivalence or an equal mix of sadness and happiness, were found to be more accurate in their final judgments and were more willing to take advice (Rees et al., 2013).

Advice taking has also been explored in individuals with psychotic-like-experiences (e.g., delusions). Scheunemann and colleagues (2020) found that while those with psychotic-like-experiences sought less advice before giving their final judgment and were more confident than individuals without psychotic-like-experiences, they did not utilize advice less than their non-psychotic-like-experiencing peers. In addition, work conducted by Kaliuzhna and colleagues (2012) found that those with schizophrenia were able to consider advice to revise their beliefs, were no more confident in their judgments than healthy controls and were more likely to place greater weight on advice than controls.

It is important to note that several papers authored by Dr. Francesca Gino were omitted from this manuscript that could shed light on the role of emotion and other applicable factors in advice taking. Recent reports that Gino has fabricated her data multiple times makes her work unreliable. Thus, papers with her as an author has been omitted. These papers include Gino and Moore (2007); Gino (2008), Gino and Schweitzer, 2008, Gino, Shang, and Croson (2009), Tost, Gino, and Larrick (2012), and Gino, Brooks, and Schweitzer (2012). For more information on Gino's activities, please see Hamid (2023) and Scheiber (2023).

The Roles of Confidence & Metacognition in Advice Taking

In psychological research, confidence judgments are often treated in one of two ways. The first is that decision-making confidence is a personality-level factor, with some individuals having a predisposition to low or high confidence regardless of their accuracy. The second is that confidence is a momentary judgment of the likelihood that one is correct (Stankov, Kleitman, & Jackson, 2015). The advice-taking literature almost exclusively treats confidence as a momentary judgment of accuracy. However, examining confidence as an individual difference will help engender the ability to reliably profile individuals' decision-making tendencies (Jackson & Kleitman, 2014), and may help explain the high variation seen in advice taking patterns (e.g., Pescetelli, Hauperich, & Yeung, 2021).

Examples of studies that treat confidence as a momentary judgment of accuracy include Wang and Du (2018), who investigated the role of trust and confidence in participants' willingness to take advice during a coin estimation task. They found that when individuals' confidence was high (M = 3.43, SD = 1.12 on a scale of 0 to 6) and their trust in the advisor was low (M = 2.22, SD = 1.40 on a scale of 0 to 6), participants tended to discount advice (WOA M = .24, SD = .22). Conversely, lower confidence and higher trust led to greater use of advice.

Sniezek and Van Swol (2001) also investigated the role of trust and an *advisor's* confidence on advice use. They found that when an advisor's confidence in their decision was higher, the judge trusted their advice to a greater extent, r = .35, p < .01, with advisor's confidence being a good predictor of their actual accuracy, r = .71, p < .0001. However, this research did not capture the judge's confidence in their initial decision, which could moderate the role between trust and use of an advisor's advice.

Additional research has supported the preposition that high decision-making confidence correlates with lower advice utilization (See et al., 2011; Pescetelli, Hauperich, & Yeung, 2011; Zhang, Harrington, & Sherf, 2022). Further work by Pescetelli, Hauperich, and Yeung (2021) sheds light on how confidence impacts advice seeking and use. Their research demonstrates that lower decision-making confidence led to greater advice seeking from advisors (M = 25% probability of seeking advice). However, advice seeking rates varied substantially from person-to-person, with probabilities of seeking advice ranging from 1% to 79%.

To examine the reasons behind this high variability in seeking advice, Pescetelli, Hauperich, and Yeung used the area under receiver operating characteristic (ROC) curves to examine participants' calibration. Calibration is how accurately an individual's confidence predicts their accuracy (Pescetelli, Hauperich, & Yeung, 2021). The area under the ROC curve (AUC) are considered a measure of internal metacognitive processes, with a value of 0.5 indicating no systematic relationship between one's confidence and their willingness to seek advice and a value of 1 indicating a perfect relationship between confidence predicting advice seeking (Fleming & Lau, 2014; Pescetelli, Hauperich, & Yeung, 2021). Participants' AUC values ranged from .41 to .97, with an average of .77. Higher AUC values correlated with a higher likelihood of seeking advice, $r = .67$, $p < .001$.

This study did not use AUC as a measure of advice use, only of the propensity to seek advice, which is important to do, as the study did find that not every time participants requested advice, did they end up using it. However, overall, when individuals specifically requested advice, they were more likely to use it. In addition, when the advice disagreed with the judge's initial opinion, they were more likely to disregard advice (M = 32% use; Range: 0% to 72%).

Additional factors were found to influence advice use in Pescetelli and colleagues' work: participants' confidence in their initial answer, initial answer accuracy, and whether advice was requested or imposed. Individuals with high pre-advice confidence tended to discount advice, but those who were initially incorrect in their answer were 10% *more* likely to use advice (35% when incorrect versus 25% when correct). The ability to request advice and when the advice matched participants' initial answer also led to higher advice use.

Furthermore, Pescetelli and colleagues' study revealed a high degree of individual variation in advice taking. Some individuals always ignored the advice, some changed their confidence without changing their final answer, and some frequently took the advice but did not change their confidence levels. The authors were unable to fully explain this phenomenon. Therefore, to better categorize this individual variation, it's critical to consider confidence as an individual difference factor that may drive this willingness or reluctance to use advice.

In addition, the researchers did not consider participants' task skill. Insights from the Dunning-Kruger effect literature highlights that task skill influences participants' awareness of their errors, with low-skilled individuals displaying a high degree of overconfidence in their abilities (Kruger & Dunning, 1999; Dunning, Johnson, & Ehrlinger, 2003; Ehrlinger et al., 2008; Dunning, 2011). Previous research has highlighted that self-perceived expertise hinders advice use (See et al., 2011; Zhang, Harrington, & Sherf, 2022), therefore overconfidence may lead to a high degree of advice discounting. In the next section, the Dunning-Kruger effect's role in shaping advice-taking will be examined.

Skill and Advice-Taking: The Dunning-Kruger Effect

A large body of literature shows that individuals are particularly poor at assessing their own abilities (Kruger & Dunning, 1999; Alicke & Govorun, 2005; Pavel, Robertson, & Harrison, 2012; Gibbs et al., 2017; Sullivan, Ragogna, & Dithurbide, 2019; Lyons et al, 2021). One of the earliest examples of self-assessment optimism is participants' rankings of their skills and abilities during the 1976 SAT Survey study. Sixty percent of one million participants considered themselves above the median in athletic ability, seventy percent in leadership, and eighty-five in cooperation with others, which is a statistical impossibility (Alicke & Govorun, 2005).

The error of misestimating one's performance is not limited to just social comparisons of one's abilities. Kruger and Dunning (1999) demonstrated that individuals also struggle to accurately assess their own baseline abilities. Kruger and Dunning (1999) observed this when completing a humor detection, logical reasoning, or English grammar task. Specifically, Kruger and Dunning (1999) found that participants in the bottom 25% of performers were grossly overconfident in their abilities (i.e., their score) and how they compared to others (i.e., percentile rankings). However, those in the top 25% showed the reverse pattern, where they underestimated their abilities and how they compared to others. The pattern of misestimation in the bottom and top performers was titled Dunning-Kruger effect by Kruger and Dunning (1999).

To explain this effect of skill on underestimation in the top quartile and overestimation in the bottom quartile, Kruger and Dunning (1999) proposed the Dual Burden hypothesis. Dunning (2011) summarized it as follows:

> "Essentially, bottom performers overestimate their proficiency because their intellectual deficits deprive them of the resources necessary to recognize that they are choosing incorrectly. They make the mistake of thinking that all their choices are at least reasonable… The problem for top performers is different. They have ample resources to know when they are most likely to be right or wrong in their choices. They get themselves right. What they get wrong is other people. Because correct answers come relatively easy to them, they mistakenly believe that other people must be coming to the same correct choices." (pp. 270-271)

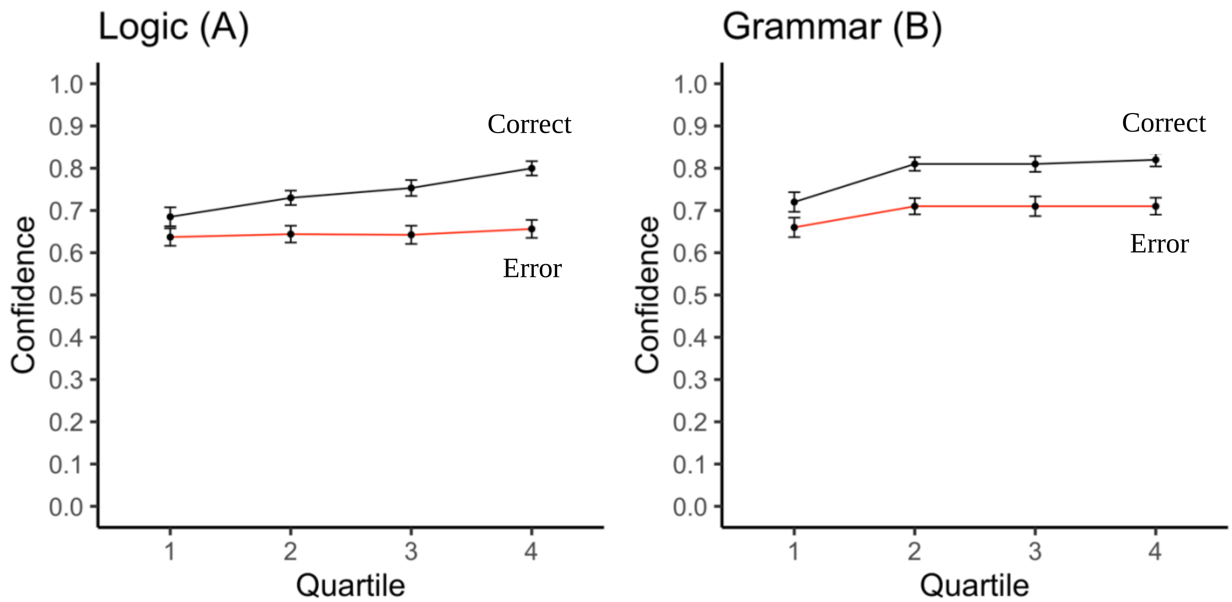
The Dual Burden hypothesis has critical implications for the likelihood that an individual will take advice and do so in an accurate manner. First, individuals with high confidence in their abilities may fail to capitalize on advice, even when it may help them correct errors. Second, errors related to misestimating other individuals' skills and abilities may lead to over reliance on advice.

Based on the tenets of the Dual Burden account, two hypotheses emerge for advice taking. Specifically, those with low skill, due to their inability to know what good performance looks like and to know when they are likely to be in error, will fail to accept advice when they should. Those with high skill, on the other hand, will utilize advice to correct erroneous decisions and ignore advice when their initial decision is correct.

However, recent work by Sanchez, Benson, and Ruthruff (2023; submitted) demonstrated that, regardless of an individuals' domain skill, they perpetrate high confidence errors (see Figure 1). Our results show that both low-skilled individuals (Quartile 1) and high-skilled individuals (Quartile 4) have nearly the same degree of overconfidence when they are wrong. This phenomenon has been titled the Universal Error Blindness Hypothesis.

Figure 1.

Average Confidence on Error and Correct Trials by Performance Quartile for a Logic (A) & a Grammar (B) Task



Note. Error bars represent 95% confidence intervals. Graph taken from Sanchez, Benson, and Ruthruff (2023).

Therefore, this research will not only analyze rates of advice use by skill groups, as per Dunning-Kruger effect convention (e.g., Kruger & Dunning, 19999), but will also explore advice taking on individuals' error trials to explore the interaction between confidence, skill, error, and advice utilization.

Personality Traits

Prior literature has largely neglected the role of individual personality and cognitive differences in advice use. Due to the lack of investigation, a strong theoretical foundation regarding the role of personality in advice taking is absent. However, existing investigations in the advice-taking and social conformity literatures have identified the roles of decision-making confidence, metacognition, and normative social influence in advice/recommendation acceptance. Drawing from the decision-making confidence (e.g., overconfidence and Dunning-Kruger effect), metacognition, and social conformity literatures, the Big-5 personality inventory (John & Srivastava, 1999) and the Susceptibility to Persuasion Scale (Modic, Anderson, & Palomäki, 2018) were identified as the most applicable source of personality factors for investigating advice utilization. These metrics will be explored in depth in the following sections.

Big-5 Personality Inventory. The Big-5 personality inventory, which is comprised of the traits Conscientiousness, Agreeableness, Neuroticism, Openness, and Extraversion, is the most frequently used personality inventory to-date to investigate personality (Feher & Vernon, 2021). See Table 1 for a definition of each of the Big-5 personality dimensions.

Table 1.

Big-5 Personality Dimensions

Personality Dimension	Meaning
Conscientiousness	Conscientiousness is how efficient and orderly an individual is and includes traits such as dutifulness, deliberativeness, low impulsivity, and thoroughness.
Agreeableness	Agreeableness is a prosocial trait, meaning it is concerned with cooperation, group unity, and social harmony. Additional dimensions include forgiving, unobtrusive, warm, modest, and giving.
Neuroticism	Neuroticism is typified by anxiety, irritability, shyness, and moodiness.
Openness to Experience	Openness to experience is concerned with curiosity, imagination, artistic expression, aesthetics, and excitability.
Extraversion	Extraversion is characterized by how sociable and outgoing an individual is and includes traits such as gregariousness, assertiveness, enthusiasm, and adventurousness.

Note. Definitions of these traits were derived from John and Srivastava (1999) and was taken from Sanchez and Speed (2020).

The Big-5 personality inventory has been used to investigate cognitive ability in older adults (Curtis et al., 2015), overconfidence in highly trained individuals (Sanchez & Speed, 2020), susceptibility to fake news (Pennycook & Rand, 2020) and occupational and academic performance (Hurtz & Donovan, 2000; Nguyen, Allen, & Fraccastoro, 2005). Despite the extensive use of Big-5 personality in investigating performance in a variety of domains, it has yet to be explored as a factor in driving decision-makers' advice taking. Only one advice taking paradigm has used the Big-5 in its investigation, but it was used to measure advisor's personality and the willingness of advisees to take financial investing based on these advisors' personality (Tauni et al., 2020). Therefore, to understand the role different Big-5 personality traits may play in advice taking, research in metacognition, overconfidence, and executive functioning was leveraged.

Conscientiousness. Conscientious individuals are detail-oriented, competent, and self-driven (John & Srivastava, 1999). Individuals with high levels of Conscientiousness have been found to have greater levels of metacognitive ability, which indicates that these individuals are better able to detect their errors than low-metacognition individuals (Kelly & Donaldson, 2016; Chiaburu, Cho, & Gardner, 2015b; Winne, 1996; Buratti, Allwood, & Kleitman, 2013; Pennycook et al., 2017; Fleming & Lau, 2014). Conscientiousness has also been implicated in individuals' ability to shift between multiple tasks, as well as working memory capacity and performance (Fleming, Heintzelman, & Bartholow, 2016; DeYoung et al., 2010; Studer-Luethi et al., 2012; Dima et al., 2015; Saylik, Szameitat, & Cheeta, 2018). The implications of this are that highly conscientious individuals are more likely to be able to rethink their initial answers than their less conscientious peers and thus may be more willing to accept advice.

Openness to Experience. Those high in Openness to Experience ("Openness") are curious, imaginative, and excitable (John & Srivastava, 1999). Prior investigations have found that individuals high in Openness were more confident in their judgments and have higher confidence on their incorrect trials than their correct trials (Buratti, Allwood,

& Kleitman, 2013). In addition, those who are more confident in their abilities are less likely to take advice than their less confident peers (e.g., See et al., 2011) has found that. Therefore, those high in Openness may be less likely to take advice.

Agreeableness. Those high in Agreeableness strive for social harmony and therefore tend to agree with those around them, even when it conflicts with their initial decision-making (John & Srivastava, 1999). Therefore, those high in this trait may be more likely to accept advice from those around them, especially from a crowd of individuals, due to their desire to maintain social harmony. However, those high in Agreeableness have been found to be overconfident in prior research, (Sukenik, Reizer, & Koslovsky, 2018), which may inhibit advice taking, though not all investigations have concurred with the link between Agreeableness and overconfidence. Prior Dunning-Kruger effect investigations have found that Agreeableness did not predict overconfidence on a logic and grammar task in highly trained individuals (Sanchez & Speed, 2020).

Neuroticism. Neuroticism is defined as a personality-level likelihood of responding to situations with negative emotions (i.e., anxiety, irritability, depression) (John & Srivastava, 1999; Lahey, 2009). Prior advice taking research has found that those who suffer from depression or emotional ambivalence were more likely to take advice than their healthy peers (Hofheinz et al., 2017; Rees et al., 2013). Therefore, those who are high in Neuroticism are hypothesized to be more likely to take advice than their low-Neuroticism peers.

Extraversion. Previous work by Kausel et al. (2015) controlled for Extraversion in their analysis examining the link between narcissistic personality disorder and advice taking, as these authors hypothesized that the positive affect and warmth dimensions of

Extraversion could promote advice taking. They found a small positive effect of Extraversion, with high extraverted individuals having higher WOA scores, $\beta = .13$, $p < .05$. Therefore, those high in Extraversion are expected to take more advice than their less extraverted peers.

Susceptibility to Persuasion. Susceptibility to persuasion measures an individual's propensity to conform to social norms of a group. Social norms are unspoken rules that govern an individual's behavior in a specific situation that are outlined by other individuals or a group (Fehr & Fischbacher, 2004; Neville et al. 2021; Cialdini & Goldstein, 2004). Social norms are thought to arise from two different influences: informational or normative influences (Deutsch & Gerard, 1955).

Informational influences are those that signal the "right" course of action, with the focus of the decision-maker being on the accuracy of the decision. When informational influence is in operation, individuals accept the group's judgment due to the opinion that the group's decision-making is superior to that of the individual (Kuan, Zhong, & Chau, 2014). Normative influences are focused on group acceptance, with individuals accepting crowd opinion due to the desire to be liked (Neville et al., 2021).

Normative-based social conformity is thought to occur due to reward learning, with successful behaviors being reinforced and errors adjusted (Klucharev et al., 2009). When behavior or judgments conflict with social norms, it is registered as an error to the decision-maker, which requires them to adjust to conform with social norms (Klucharev et al., 2009; Cialdini & Goldstein, 2004). As Cialdini and Goldstein (2004) remark, conformity garners social approval, which not only is rewarded with relationship building, but can also help bolster one's self-esteem. Violation to the socially presented option may be met with punishment in the form of ostracization by the group (Hertz &

Wiese, 2018). Therefore, individuals will go with the group consensus even when it leads to an incorrect choice, to be accepted within the group. As such, I hypothesize that an individuals' propensity for normative social influence will lead to greater advice utilization, as individuals will accept advice offered due to the desire to be accepted by other individuals.

Chapter II Summary

The advice-taking literature has an extensive history of exploring the advisor and judge behaviors that lead to greater advice utilization. Despite this rich history, two fundamental research gaps remain. First, the effect of multiple advisors on an individuals' willingness to accept advice is unclear. Prior research in the social conformity area has consistently found a positive relationship between number of advisors and recommendation adherence. However, due to large methodological differences (e.g., timing of recommendations/advice), it is unclear whether the effect will replicate in advice taking domains when individuals have already cognitively committed to their answer.

The one advice-taking study that did examine the effect of multiple advisors on advice utilization failed to find this positive relationship, with participants taking less advice from nine advisors than they did from two or four (Mannes, 2009). However, the authors did not control the accuracy of the advice, therefore inaccurate advice was likely included in what was presented to participants. Accuracy of the advice is a critical moderating factor, as individuals have been shown to be able to detect the general accuracy of the advisor (e.g., Harvey, Harries, & Fischer, 2000; Yaniv & Kleinberger, 2000) and are less willing to take advice from advisors they perceive to have lower expertise (Bailey et al., 2022). Therefore, the research presented in this dissertation will provide high-quality advice from a single and group advisor to better explore the link between multiple advisors and advice taking.

The second identified gap in current advice-taking research is the lack of knowledge regarding individual differences in advice taking. Current studies that have examined judge-specific factors on advice taking have focused on one of two areas, the role of state-based factors or the role of clinical disorders in advice taking. State-based factors, such as emotion, self-esteem, "state-level narcissism," and perceived personal power, have been found to be an inhibiting factor on advice taking. On the other hand, clinical disorders, such as depression and schizophrenia, have been found to lead to greater advice taking. However, the roles of skill, metacognition, and personality remain relatively unexplored.

Research in the Dunning-Kruger effect field highlights the fact that skill and an individual's ability to know when they are likely to be incorrect (i.e., metacognitive sensitivity) may be key factors in promoting advice taking. Skill and metacognitive sensitivity are thought to interact with one another to enable a need-for-advice judgment, with high skill expected to enhance individuals' metacognitive ability (Dunning, 2011). Therefore, those with high skill are thought to know when they have made an error and will seek advice, while those with low skill are thought to be blind to their errors and thus less willing to seek advice (the Dual Burden Account; Dunning, 2011). However, recent research has highlighted that this metacognitive deficit on one's high-confidence error trials may be shared by both the low- and high-skilled, which may inhibit advice taking by either group when in a high-confidence state (Sanchez, Benson, & Ruthruff, 2023). Therefore, the current work explores the link between skill, metacognition, and advice taking as a skill-based factor or as a confidence-based factors.

As for personality, a strong theoretical foundation regarding the role of personality in advice taking did not exist prior to this investigation. Therefore, research in the metacognition, overconfidence and Dunning-Kruger effect, and social conformity literatures highlighted the potential roles of the Big-5 personality characteristics and normative social influence in advice taking. The traits of *Conscientiousness*, *Agreeableness*, *Neuroticism*, *Extraversion,* and *Normative Social Influence* are thought to lead to enhanced advice taking due to the links of these traits with the ability to rethink one's answer (Conscientiousness), the desire to maintain positive social relationships. (Agreeableness, Extraversion, and Normative social Influence), and provide realistic judgments of one's performance (Neuroticism). Those high in Openness are thought to take less advice due to their propensity for overconfidence. The next chapter will highlight specific hypotheses and predictions that have emerged from this critical review of the current advice taking literature in detail.

Chapter III: Research Aims & Predictions

Three primary gaps in advice-taking literature have been identified. First, this research addresses whether advice coming from a group of advisors will lead to greater advice utilization than that of a single advisor due to normative social influence. The second gap this research fills is examining the role of metacognitive error monitoring in advice utilization, with a specific examination of the role skill plays in enhancing or inhibiting advice usage. Third, although past research suggests large individual differences, we as a field are unable to accurately predict who is most likely to accept advice. Therefore, the final aim of this research is to empirically test identified personality characteristics that are hypothesized to lead to high advice utilization or discounting. To address these gaps, this work explores the following hypotheses, which are not mutually exclusive, but may work in tandem with one another to enhance or inhibit advice taking:

Characteristics of the Advisor Hypotheses

1. **Hypothesis 1**: *The Social Consensus Hypothesis* – Due to the normative and informational social influence a crowd exerts, individuals are more likely to believe in the crowd advisor and use their advice, compared to a lone advisor.
    a. Prediction 1: Individuals will have higher ratings of trust for the advice coming from the crowd advisor than that of the single advisor.
    b. Prediction 2: Individuals will rate the crowd as being more accurate than the single advisor.
    c. Prediction 3: Individuals will use the crowd's advice more than the single advisor's advice.

Empirical results of these hypotheses will be covered in *Chapter V: Advice Taking Results.*

Metacognition, Confidence, and The Dunning Kruger Effect Hypotheses

1. **Hypothesis 1**: *The High Confidence Leads to Advice Blindness Hypothesis* – People are especially sensitive to their internal subjective feelings of confidence in their initial answer, and relatively insensitive to external cues, such as disagreement between their and the advisor(s)' answer.
    a. Prediction 1: Previous literature has shown that when individuals are in a high confidence state, they are less likely to seek or accept advice (e.g., Kausel et al., 2015; De Wit et al., 2017). Therefore, a negative relationship is expected between high initial confidence and advice use.
    b. Prediction 2: There will be little or no difference in advice taking rates amongst the low-skilled and high-skilled on their high confidence error trials, as their confidence in their initial answer blinds them to the need for advice.
2. **Hypothesis 2**: *The Skill-Based Advice Taking Hypothesis* – This hypothesis is based on the Dual Burden hypothesis of Dunning (2011) applied to the advice-taking domain. The high-skilled, due to their greater ability to identify when they are incorrect and their superior domain skill, are better able to capitalize on the advice to rethink their initial answer.
    a. Prediction 1: Thus, advice taking on both low and high confidence error trials will be higher in the high-skilled than the low-skilled.

Results of these hypotheses will be covered in *Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results.*

Individual Difference Hypotheses

3. **Hypothesis 5**: The Big-5 Hypotheses:
    a. **Hypothesis 5.a.:** Conscientiousness has been implicated in individuals' ability to shift between multiple tasks and working memory capacity/performance (Fleming, Heintzelman, & Bartholow, 2016; DeYoung et al., 2010; Studer-Luethi et al., 2012; Dima et al., 2015; Saylik, Szameitat, & Cheeta, 2018), which enables high Conscientiousness individuals to rethink their initial answer to catch their errors.
        i. Prediction 1: Highly Conscientious individuals will take more advice.
    b. **Hypothesis 5.b.:** Individuals high in Openness are more confident when judging their initial performance and often fail to revise their judgments due to their expectation that they have performed well in a domain (Buratti, Allwood, & Kleitman, 2013). Therefore, those high in Openness are expected to be less likely to take advice due to their erroneous expectation of good performance, which blinds them to the need for advice.
        i. Prediction 1: A negative relationship will exist between Openness and advice use.
    c. **Hypothesis 5.c.:** Those who are high in Agreeableness are likely to accept advice due to their desire to acquiesce to others, regardless of whether the advice comes from a crowd or from a single advisor.

i. Prediction 1: A positive relationship will exist between Agreeableness and advice use.

d. **Hypothesis 5.d.:** Those who are high in Neuroticism will be *more* likely to accept advice due to "depressive realism" (Alloy & Abramson, 1988), which is the phenomenon that individuals who are depressed are more accurately able to estimate their abilities.

i. Prediction 1: A positive relationship will exist between Neuroticism and advice use.

e. **Hypothesis 5.e.:** Individuals high in Extraversion are likely to accept advice due to their positive affect and warmth, which leads them to see the best in others and desire to maintain positive social interactions (Kausel et al., 2015).

i. Prediction 1: A positive relationship will exist between Extraversion and advice utilization.

2. **Hypothesis 6**: Those high in normative social influence are more likely to accept advice due to their desire to belong to the group consensus.

a. Prediction 1: A positive relationship will exist between normative social influence and advice taking.

b. Prediction 2: Due to the normative influence a crowd exerts on an individual, those higher in susceptibility to normative influence will take more advice from the crowd then the advisor, whereas those lower in susceptibility will not.

i. Therefore, an interaction between advisor type and normative social influence is expected.

Results of this analysis will be covered in *Chapter VII: Individual Difference Investigation Results*.

Chapter IV: Method

Participants & Recruitment

All work conducted was approved by UNM's institutional review board under protocol 2301035339A002. Participants were restricted to those who were 18 years of age or older and were located in the United States. Participants were paid $10 for one hour of their time.

This study recruited participants from recruitment platform Prolific. Prior research conducted by Douglas, Ewell, and Brauer (2023) has indicated that compared to undergraduate samples collected through SONA, Qualtrics recruiting, and Amazon Mechanical Turk (MTurk), Prolific offers superior data quality. Participants on Prolific were found to pass attention checks to a greater degree, provide meaningful answers in open-text responses, follow instructions, work slowly enough to understand all of the material, and were more likely to remember previously presented information during different phases of experimentation.

Exclusion Criteria

244 participants took part in this study. Two metrics were used to determine whether a participant should be excluded for not following study protocols. First, participants were removed if they did not complete the experiment in its entirety. This led the removal of seven participants. Second, participants were removed if they had indicated on a question at the end of the experiment that we should remove their data due to guessing randomly, rushing through the study, or using outside resources to answer the logic questions. This led to the removal of two additional participants.

Finally, 11 participants had achieved 100% accuracy on the logic questions in either the single advisor condition or the crowd conditions. As the key metric of interest is the rates at which individuals accepted advice on their error trials from both the single advisor and the crowd, these participants were excluded from analysis. This left a total of 224 participants in the final sample.

Materials

Demographics and the Personality Battery

Participants were asked to provide their age, their gender assigned at birth, and their current gender identity. After completing the demographic questions, participants completed two measures of personality. Table 2 lists the personality measures used in this study, the purpose of each measure, and the reliability of each measure. For each of the personality batteries, the order of the questions was randomized to prevent item order effects. Following the personality battery, participants were administered a 26-item logic task.

Table 2.

Measures of Personality and Cognition

Measure	Items	Purpose	Reliability
NEO-PI-3 (John & Srivastava, 1999)	44-items	Measures the Big-Five personality traits, which include the dimensions of: 1. ***Conscientiousness:*** organized, thorough, deliberate 2. ***Agreeableness:*** trusting, forgiving, acquiescing 3. ***Neuroticism:*** shy, moody, irritable 4. ***Openness to Experience ("Openness"):*** curious, imaginative, artistic 5. ***Extraversion:*** outgoing, energetic, sociable	α = .88 to .92 (McCrae, Costa, & Martin, 2005)
Susceptibility to Persuasion Scale – II (STPS-II) (Modic, Anderson, & Palomäki, 2018)[a]	27-items	Measures an individual's susceptibility to persuasion from other individuals, which includes the dimensions of: 1. **Premeditation:** An individual's ability to foresee and mitigate the consequences of a course of action. 2. **Consistency:** the need to remain consistent in one's thoughts and behaviors in front of others. 3. **Sensation Seeking:** an individual's tendency to seek out new experiences, sometimes at the risk of their wellbeing. 4. **Self-Control:** an individual's ability to control their thoughts and behaviors in the face of tempting or aversive stimuli. 5. **Normative Social Influence:** the likelihood that an individual will conform to a group's ideologies or actions. 6. **Similarity:** An individual's desire to stand out from a group, to be unique. 7. **Risk preference:** an individual's comfort with financial risk-taking. 8. **Advertising attitudes:** An individual's acceptance of advertisements. 9. **Need for cognition:** one's propensity to engage in intellectual activities. 10. **Uniqueness:** the desire to have choices, rather than run-of-the-mill options.	α = .75 to .92

[a] The short-form was used

The Logic Task

The logical reasoning task was composed of several well-established logical reasoning batteries. In total, 11 questions from the Frederick (2005), Thomson and Oppenheimer (2016), and Toplak, West, and Stanovich (2014) Cognitive Reflection Tests (CRT) were used in conjunction with 13 questions from the Heuristics & Biases battery (Toplak, 2011). Two additional probability questions from Tversky and Kahneman (1974) were combined with the aforementioned questions.

These questions were chosen as the stimuli for this experiment due to their problem-solving nature and their difficulty. To measure research outcomes, it was necessary to have questions for which participants were likely to make errors. Tasks for which participants are too proficient will yield little to no advice taking because participants' accuracy is at ceiling, while tasks participants have no knowledge of will yield over-reliance on the advice due to floor effects. These logic questions have been used in a prior study (N = 268; Sanchez, Benson, & Ruthruff, 2023), enabling the calculation of the average difficulty of each question. The mean task score in Sanchez, Benson, and Ruthruff (2023) was M = 11.05 questions (SD = 4.21 questions).

In addition, as we wanted to determine whether individuals' faulty reasoning could be corrected by a nudge produced through advice, it was important to have questions that involved a procedural process like math or logic rather than a binary "I know" or "I do not know" based problem, like general trivia questions.

Performance Estimates

For the Logic task, participants were asked to provide pre- and post-estimates of their performance, as well as provide item-by-item accuracy estimates. The pre- and post-estimates asked participants to 1) estimate what percentage of the task they would/did answer correctly, 2) estimate what percentage of the task *others* would/did answer correctly, and 3) to estimate out of 100 other participants completing the task, how many participants they believed they would/did score higher than.

The Advice

Participants were told that for half of the questions they would receive advice from a group of prior participants completing the task (crowd manipulation) and on the other half they would receive advice a prior student who had completed the task (single advisor manipulation). The advisor manipulation was similar to prior studies, in that participants were told it was coming from a peer who had also completed the task (e.g., Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; Soll & Larrick, 2009; See et al., 2011).

For each trial, participants first provided an answer to each logic question and a confidence judgment regarding their certainty in their accuracy. Following this, participants were re-shown the question along with advice from the crowd or advisor. Advice was presented in the form of a histogram. This graph highlighted either the frequency for which each answer option was chosen (i.e., the crowd condition) or ratings of likelihood that each answer option was correct (i.e., for the lone advisor condition). To emphasize which answer the advisor or crowd was advocating for, a statement was provided to the participant stating that "Based on this distribution, the [advisor or crowd] has indicated that [*XYZ answer*] is correct." Examples of the advice shown to participants are shown in Figure 2 and Figure 3. It is important to note that the distribution graphs used for the advisor and crowd manipulations were exactly the same across conditions. Only the stated source of the advice was varied across conditions.

Figure 2.

Advice from a Prior Participant Advisor

Question

In her introductory psychology class, Anna received both the 14th highest and the 14th lowest test score in the class. How many students are in the class?

Prior Participant's Answer

Below you will find information regarding how a prior participant answered this question. The participant was asked to rate the probability that each answer option is the correct answer. Based on this distribution, the prior participant has indicated that *27* is the correct answer.

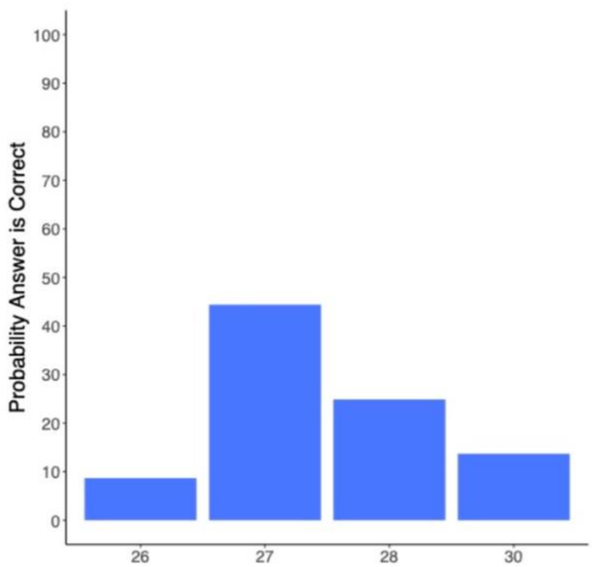
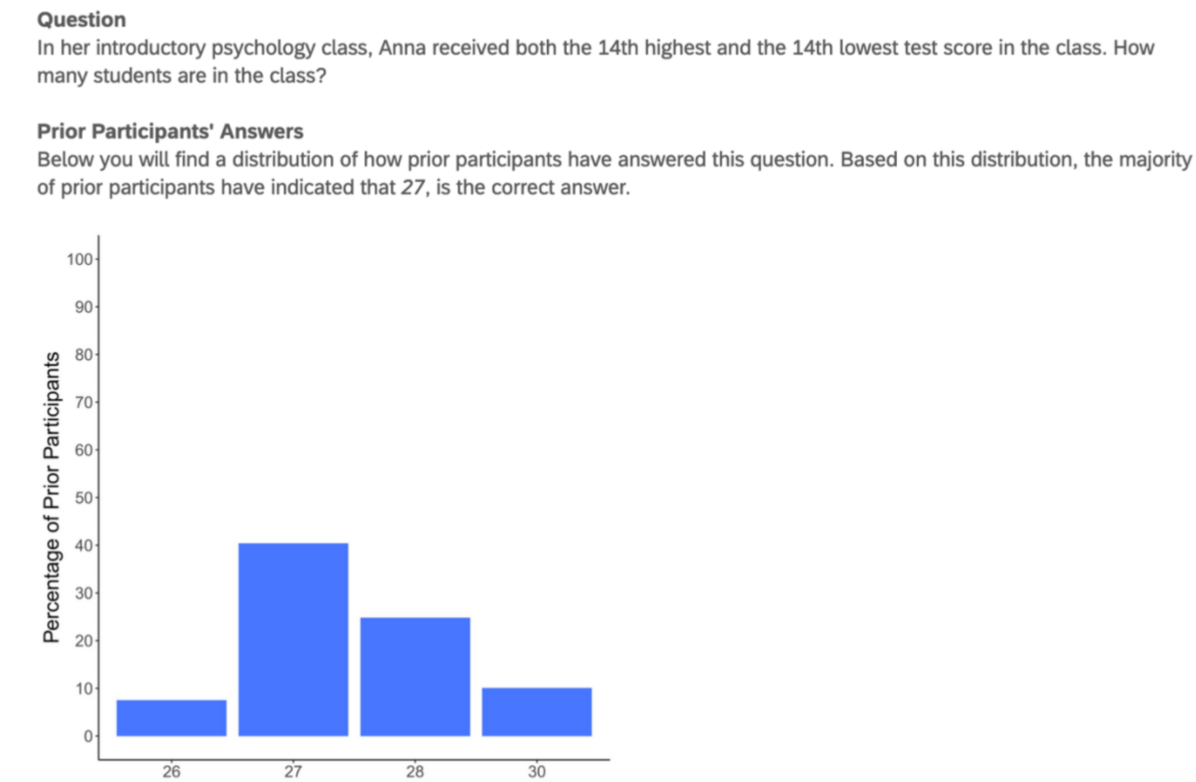


Figure 3.

Advice from a Group of Prior Participants



To generate the distribution graphs, we began with actual data from Sanchez, Benson, and Ruthruff (2023). Using R's ggplot package (Wickham, 2016), the data were altered somewhat, so that the highest bar was always 20-25% higher than the second highest bar. Furthermore, we also made the most popular answer the correct one. The second choice was always the top "foil" – the incorrect answer that actual participants chose most often (and thus the answer frequently chosen by the present participants).

Procedure

Participants first provided their consent electronically by selecting "I CONSENT". If they did not consent, then they were exited from the survey. After consenting to participate, they were provided a short description of the structure of the experiment with estimates of how long each section would take to complete. As part of the instruction screen, participants were encouraged to answer the questions carefully.

To help ensure good data quality, participants were asked prior to starting the task to agree not to use outside resources (e.g., the internet or other individuals) to answer the research questions. A market research study conducted by Geisen (2022) has found that when participants are asked to commit to providing quality answers, data quality issues decreased by 50% compared to the control group. Should participants indicate that they did not agree, they were exited from the experiment.

After providing their commitment to not use outside resources, participants were given the demographic and personality questions. Following this, participants were asked to provide pre-assessments regarding their anticipated performance on logic questions. After completing the pre-assessments, they were transitioned to the logic task.

The Logic Task. The presentation of the logic questions and the advisor and crowd conditions were counterbalanced to prevent order effects. Counterbalancing was done using a list-based approach where the logic questions were divided into two 13-item lists (A and B). The counterbalancing structure is illustrated below:

1. Advisor manipulation paired with List A[2] | Crowd manipulation paired with List B
2. Advisor manipulation paired with List B | Crowd manipulation paired with List A
3. Crowd manipulation paired with List A | Advisor manipulation paired with List B
4. Crowd manipulation paired with List B | Advisor manipulation paired with List A

[2] List A contained logic question one through number 13, while List B refers to logic questions 14 through 26.

Within List A and List B, the order of the logic questions was randomized for each participant. For each list, participants saw advice from only one of the two advisors.

During the logic task, participants were asked to first answer the logic question and then provide a confidence judgment. Following this, participants received the advice. Participants were able to view the question-and-answer options the whole time they were reviewing the advice. This was done so that participants were afforded the ability to reconsider their initial answer, not just merely accept or reject the advice. After viewing the advice, participants were asked to click the next button to transition them to a series of questions about their initial judgment and their desire to change their answer.

Deciding Whether to Accept the Advice. After seeing the advice, participants were first asked to provide a confidence judgment about their *pre-advice answer* (i.e., the answer they chose before seeing the advice). This performance judgment captured changes in participants' certainty regarding their pre-advice answer. After this, participants were asked whether they would like to change their answer, which had a yes/no answer option. When participants indicated "yes," they were re-provided the logic question and allowed to select a new answer. After providing their final answer, participants gave a confidence judgment for their final answer.

Interactions with the Advisor Survey. At the end of each logic question list (List A and List B), participants were transitioned to the advisor rating questionnaire. Participants were asked questions about their interactions with the source of the advice (i.e., crowd or single advisor). First, they were asked how often the advice given by either the majority of participants (crowd manipulation) or a prior participant (single advisor manipulation) was correct on a scale of 0% - 100%, with 0% meaning "never correct" to 100% meaning "always correct." Next, they were asked how often they used the advice offered by the majority of participants or a prior participant on a scale of 0% - 100%, with 0% meaning "I never switched to the participant's/majority of prior participants' answers" to 100% meaning "I always switched to the participant's/majority of prior participants' answers." Finally, participants were asked on a scale of 1 (not willing at all) to 5 (completely willing), how willing they were to use the answers chosen by the majority of prior participants or the prior participant. These questions were asked at two points during the experiment, once after the first 13 logic questions and once again after last 13 logic questions.

Post-Logic Task Assessments. At the end of the logic task (all 26 questions), participants were asked to provide post-assessments regarding the percentage of questions they believe they had answered correctly, the percentage of questions they believed others would answer correctly, and a percentile ranking estimate. To create their percentile ranking estimates, participants were asked "[i]magine that there are 100 other participants completing this task. Of these 100 participants, how many do you think you will score higher than?"

Exclusion Question & Debriefing. After providing their post-task performance estimates, participants were asked "[a]fter completing this experiment, is there any reason we should exclude your data from our analyses?" If participants indicated "yes" to this question, then they were presented with a second multiple choice question that asked them to specify a reason why their data should be excluded. Reasons for omission included guessing randomly, rushing, using outside resources like Google, not feeling their answers were good enough to be useful, and other. If participants indicated they did not feel their answers were good enough to be useful, then their data was kept for analysis. If participants indicated that we should omit their data for guessing, rushing, using outside resources, or another reason, then it was removed from analysis.

At the very end of the experiment, participants were presented with a debrief statement. Participants were debriefed as they were not informed prior to starting the experiment that we had simulated the histograms to create 100% accurate advice. The debrief statement is provided below.

> **Debriefing:** The histogram graphs displayed did not perfectly correspond to data from previous experiments. The data used in the histograms was based on answers provided by prior participants who had completed these questions but was altered to assess the influence on a person's decision making when the information provided was highly accurate. Please **do not share** this information with others who may take the survey as it will bias our results. Thank you so much for helping us better understand how people make decisions. We really appreciate your time. Please reach out to [our team] should you have questions or if you did not receive credit for completing this experiment. Thanks again!

Chapter V: Advice Taking Results

This chapter investigates how the source of the advice (a single or crowd advisor) may be a remedy to the problem of low-advice utilization. Across the advice taking literature, advice taking rates are consistently as low (20% to 30%), even when advisors demonstrate high accuracy levels (Harry & Fischer, 1997; Soll & Larrick, 1999; Soll & Larrick, 2009; Duan, Gu, & Sun, 2016; Pescetelli, Hauperich, & Yeung, 2021). However, social conformity research consistently finds that the more individuals there are providing advice, the more likely individuals are to take it (Farjam & Loxbo, 2023; Bond & Smith, 1996; Asch, 1956; Klucharev et al., 2009; Perfumi et al., 2019; Wijenayake et al. 2020).

Prior to the current investigation, the findings in the social conformity literature could not be directly compared to those in the advice taking literature. This is due to the different timing of advice presentations in these literatures. The advice taking literature typically presents the advice after participants have made an initial decision, while the social conformity literature presents the advice at the same time participants are presented with a stimulus. Presented advice simultaneously with the question is thought to lead to heuristic decision-making and greater advice acceptance, while advice presented after participants have made an initial decision will lead to greater resistance or "anchoring" to their initial opinion. This work is one of the only studies to directly compare the effect of a single versus multiple advisors.

The main hypothesis this chapter tests is the *Social Consensus* hypothesis, which states that due to the normative and informational social influence a crowd exerts, individuals are more likely to believe in the crowd advisor and use their advice, compared to a lone advisor. This hypothesis predicts that participants will perceive the crowd advisor as more accurate, trust the crowd advisor more, and will ultimately use more advice from the crowd advisor than from the single advisor. Should the Social Consensus hypothesis prove to be empirically true, design of advice recommendation systems can leverage social elements to enhance advice taking.

Data Processing & Analysis

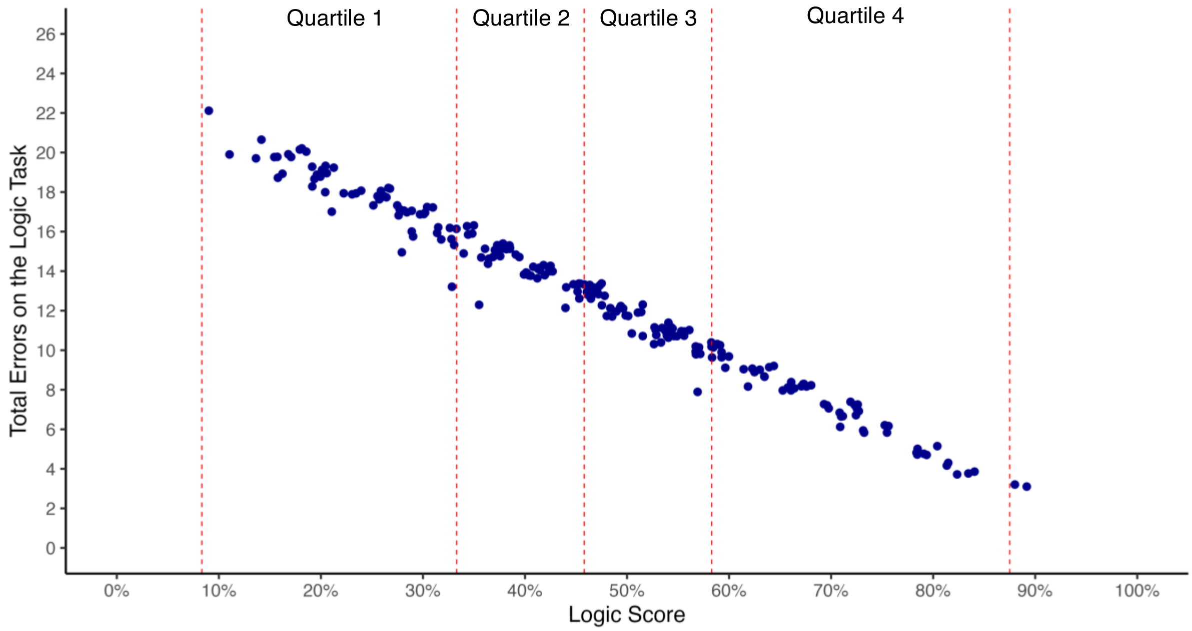
On rare occasions (0.5% of trials), participants changed their initial answer following the advice but not to the advisor's recommended answer. These trials are omitted from all analyses in this manuscript, as it is ambiguous whether participants took the advice or not. In addition, Logic trials 23 and 25 were omitted due to a technical error that led to a failure to trigger the advice. Analyses were conducted using R version 3.6.3 and the R interface Jamovi ("The jamovi project," 2023).

Advice Taking on Error Trials

First, the number of errors participants made on their error trials were assessed, as this has a direct bearing on the number of opportunities they would have to accept advice. The distribution of errors on the logic task (comprised of 24 questions) are highlighted in Figure 4. For ease of interpretation, participants were divided into four "score quartiles" based on their pre-advice task score, with those in Quartile 1 scoring in the bottom 25% of participants and those in Quartile 4 scoring in the top 25% of participants. This quartile delineation will be used to visualize advice taking and self-estimates of performance for ease of interpretation, while the continuous variable of "logic score" is used in analyses.

Figure 4.

Distribution of Errors for Each Score Quartile



Note. On rare occasions, participants changed their initial answer following the advice but not to the advisor's recommended answer. These trials were omitted. Thus, some individuals have the same logic score but not the same number of errors considered.

On average, individuals in the bottom quartile answered 17.75 questions incorrectly (*SD* = 1.69), while those in the top quartile answered 6.73 questions incorrectly (*SD* = 1.71) out of 24 logic problems. The second quartile answered 13.83 questions incorrectly on average (*SD* = .91), while the third quartile answered 10.80 questions incorrectly on average (*SD* = .90).

Next, individuals' overall probability of accepting advice on their error trials was assessed. This is an important pre-assessment before examining the role the type of advisor plays to compare advice taking patterns seen in this study to that of prior advice taking investigations in the literature. To assess advice taking patterns for the sample overall, while also considering the role of individual variability, a multi-level logistic model was used to assess each individual's (*N* = 224) probability of accepting advice on their error trials. This analysis was limited to error trials, given the 100% correct advice, as individuals could only realistically take advice on their error trials. The outcome variable accepting advice was a binary variable (0 = did not take advice, 1 = took advice), while participants' ID was entered as a random intercept variable. No other predictors were entered into the model. The equation for this model is represented below:

$$\boldsymbol{p(Switching\ to\ Advisor's\ Answer)_{ij}} = \beta_{0j}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

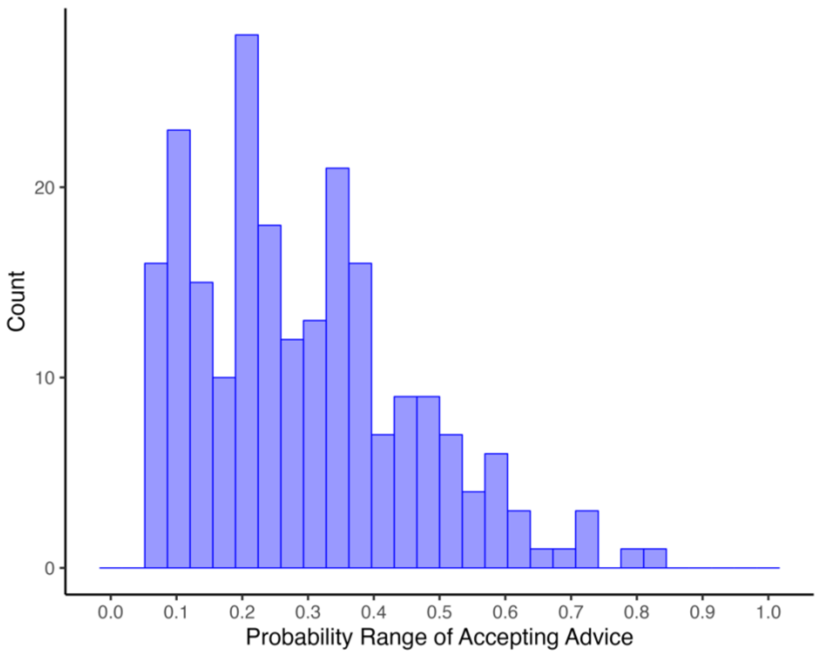
where *i* represents each logic question nested within participant *j.* β_{0j} represents the random intercept for each participant. The model was fit using maximum likelihood (Laplace Approximation) using the lme4 package's glmer() function (Bates et al., 2015).

The fixed intercept for this random intercept only model was – 1.07 on a logit scale, which represents the average log-odds of accepting advice across participants. For ease of interpretation, the fixed intercept was converted from a logit scale to a probability scale using a function provided by Van Horn (2023), which converts log-odds to a probability scale. This conversion indicated that individuals had a 25.6%, 95% CI [.22 – .30] probability of accepting advice on average, which is comparable to prior research (with advice taking rates typically ranging from 20% to 30%). However, the random intercepts have standard deviation of 1.09 log-odds, which indicates a high degree of variability across individuals in their willingness to accept advice.

To analyze this individual variation in the probability of accepting advice, the random effects were extracted from the logistic multi-level model and the fixed intercept of – 1.07 log-odds was added to each individuals' intercept standard deviation. The result of this analysis is plotted in Figure 5. Overall, despite high variability in the willingness to accept advice, the trend seen is that individuals have a relatively low probability of accepting advice.

Figure 5.

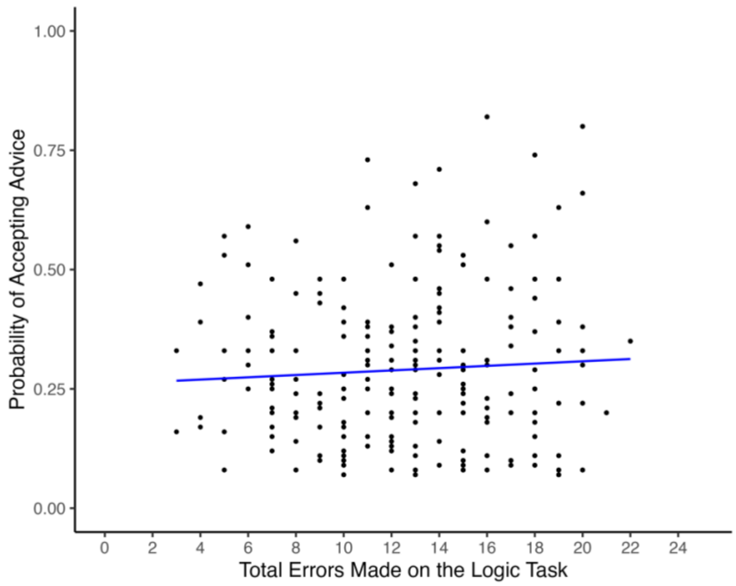
Each Individuals' Probability of Accepting Advice



As participants' opportunities to accept advice are directly tied to the number of errors they made on the task, participants' probability of accepting advice was plotted against the number of errors they made, as illustrated in Figure 6. Almost no trend existed between the number of errors participants made on the task and their probability of accepting advice.

Figure 6.

Predicted Probability of Accepting Advice for Plotted Against Total Number of Errors Each Participant Made



Advice Use from the Single Advisor Versus Crowd Advisor

Next, advice usage was examined by whether the advice came from a single advisor or from the crowd. According to the social consensus hypothesis, greater utilization will be seen from the crowd advisor than the single advisor.

Multi-Level Analysis Overview

Two multi-level models were constructed to examine the probability of accepting advice on one's error trials from the crowd versus single advisor. Models were fit using maximum likelihood (Laplace Approximation) using the lme4 package's glmer() function (Bates et al., 2015).

The first model was a random intercept only model with the advisor type (crowd or single advisor) and question difficulty entered as fixed effects. This model addresses the question of whether individuals use the crowd advisor more than the single advisor, while considering individual variability in advice taking patterns and item difficulty.

Prior literature has indicated that advice taking increases when problems are perceived as more difficult (Gino & Moore, 2007). However, due to reported misconduct by the most prolific author of this work (see Hamid, 2023), it is unclear what role difficulty may play in increasing advice taking. In the present analysis, the measure of item difficulty was calculated based on the accuracy of answers from a previous sample of college participants who had completed the logic questions (Sanchez, Benson, & Ruthruff, 2023). Difficulty was grand mean centered, with negative values indicating a difficult problem (i.e., low accuracy) and positive values indicating easy problems (i.e., high accuracy).

The second multi-level model had the same overall structure as the random intercept only model, with random intercepts for each participant and fixed effects for advisor type and item difficulty. An addition to this model was random slopes for advisor type. This allowed for the examination of whether there were individual differences in participants' use of one advisor over the other; if so, it would be possible to enhance advice taking by determining which type of advisor participants prefer and then using that knowledge to tailor advice. This final model is represented by the following equation:

$$\boldsymbol{p(Switching\ to\ Advisor's\ Answer)_{ij}} = \beta_{0j} + \beta_1 Difficulty_{ij} + \beta_{2j} Advisor\ Type\ on\ Trial_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{2j} = \beta_2 + u_{1j}$$

where *i* represents each logic question nested within participant *j.* β_{0j} represents the random intercept for each participant, while β_{2j} represents the random slope for advisor type.

According to Sommet and Morselli (2017), interpretation of fixed and random intercepts and slopes are as follows. A fixed intercept corresponds to the overall log-odds of accepting advice for the sample. Therefore, a fixed intercept assesses participants' willingness to accept advice when the single advisor is used and when the item difficulty is at the grand mean level. A fixed slope is the average change in log-odds of accepting advice when the crowd advisor is used rather than the single advisor. The random slope is the variation in log-odds for each participant when the crowd advisor is used over the single advisor, in other words, the random slope identifies individual differences in the effects of advisor type on taking advice.

Model Selection. The results of the random intercept only and random intercept and slope models are highlighted in Table 3. Akaike Information Criterion (AIC) values and Chi-squared tests were used to examine the fit of the two models. AIC estimates the level of prediction error in a model, with an AIC value of at least 2 lower considered the better fitting model (Wagenmakers & Farrell, 2004). In addition, a Chi-squared test compares the log-likelihood ratio statistics for each model to aid in model selection (Busemeyer & Wang, 2000). Model fit statistics indicated that the random intercepts and slope model fit the data best (AIC: 3091.8) in comparison to the random intercept only model (AIC: 3095.6). Chi-squared statistics between the random intercept only and the random intercept and slope models indicate that the random intercept and slope model fit the data better, $\chi^2(2, N = 224) = 7.80$, $p = .022$. Therefore, the model random slope and intercept model was selected.

Table 3.

Multi-level Model Examination of Advisor Type and Item Difficulty on the Probability of Accepting Advice

	Random Intercept Only Model			Random Intercept & Slope Model		
Predictors	*Log-Odds*	*std. Error*	*p*	*Log-Odds*	*std. Error*	*p*
Intercept	-1.06	0.10	**<0.001**	-1.09	0.11	**<0.001**
Advisor Type: Crowd	0.03	0.09	0.726	0.05	0.12	0.652
Difficulty[a]	0.19	0.20	0.335	0.19	0.20	0.346
Random Effects						
τ_{00} (Variance of Random Intercepts)				1.48		
τ_{11} (*SD* of Random Slope)				0.73		
ρ_{01} (Correlation: Intercepts & Slopes)				-0.39		
N				224		
Number of Trials				2753		

Note.
[a] Grand Mean Centered Variable

Do Individuals Use the Crowd Advisor More Than the Single Advisor & Is Advisor Preference an Individual-Level Difference?

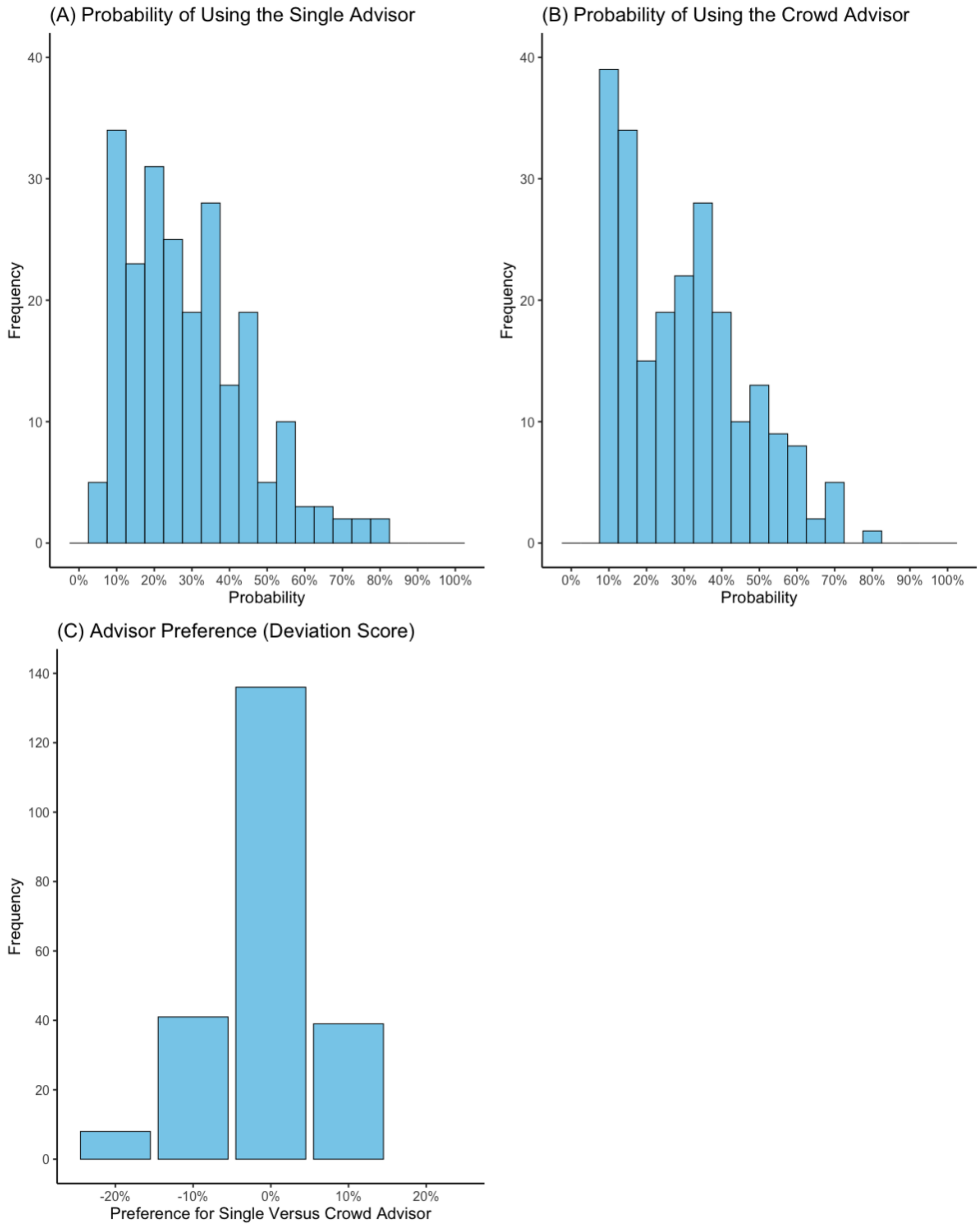
Examining the fixed effects, participants did not use the crowd advisor significantly more than the single advisor, b = .05, exp(β_2 = OR = 1.05, 95% CI [0.84 – 1.32], p = .652. In addition, the difficulty of the logic problem did not lead to greater advice use, b = .19, exp(β_1 = OR = 1.21, 95% CI [0.81 – 1.81], p = 0.346.

The fixed intercept was -1.09 log-odds (25.2% probability) of accepting advice, with random effects indicating a standard deviation of 1.21 log-odds of accepting advice across participants. This high individual variability in the likelihood of accepting advice is consistent with the high variability seen in earlier analyses in this chapter. An estimated marginal effects analysis indicates that individuals had a 24.8% probability of using the single advisor, 95% CI [.21 – .29], while participants had a 25.7% probability of using the crowd advisor, 95% CI [.22 – .30].

The random slopes, which addressed the question of whether each individual is more likely to take advice from a specific advisor would be seen at the individual level, indicated that there was significant variation amongst participants from the fixed slope of .05 log-odds, *SD* = 0.73 log-odds. The probability of accepting advice from the crowd versus the single advisor is plotted in Figure 7, alongside participants' preference for one advisor type over the other (as represented by a deviation score). Panels (A) and (B) display each individuals' probability of taking advice from the single advisor and the crowd advisor, respectively. Panel (C) demonstrates their preference for one advisor over the other, with positive values indicating a preference for the crowd advisor and negative values indicating a preference for the single advisor.

Figure 7.

Participants' Use & Preference for the Single Versus Crowd Advisor



Note. Panel (C) represents deviation scores, with 0 indicating no difference in preference for the single or crowd advisors. A negative score indicates a preference for the single advisor, while a positive score indicates a preference for the crowd advisor.

Panel (C) was created by taking each individuals probability of accepting advice from the single advisor (i.e., the random intercepts) and their likelihood of accepting advice from the crowd advisor (i.e., the random slopes) and calculating the deviance between these probabilities. First, participants' random intercepts and slopes were converted from log-odds units to probabilities. Then these probabilities were subtracted from one other. This process is illustrated by the following equation:

$$\frac{exp(fixed\ intercept + random\ intercepts)}{1 + exp(fixed\ intercept + random\ intercepts)} -$$

$$\frac{exp(fixed\ intercept\ +\ (slope\ coefficient\ +\ (random\ intercept\ +\ slopes))}{1 +\ exp(fixed\ intercept\ +\ (slope\ coefficient\ +\ (random\ intercept\ +\ slopes))}$$

Negative deviation scores indicates that participants prefer to use the single advisor over the crowd advisor. Positive deviation scores indicate that participants prefer the crowd advisor over the single advisor. Deviations of zero indicate equal probability of using either advisor. Original code based on the above equation was provided by Van Horn (2023) and was modified to conduct this analysis.

Approximately 61% (N = 136) of the sample had deviation scores of zero, indicating no preference for either advisor. Approximately 18% (N = 41) of participants were 10% more likely to use the single advisor over the crowd advisor, while approximately 4% (N = 8) of participants were 20% more likely to use the single advisor than the crowd advisor. Approximately 17% demonstrated a 10% greater likelihood of using the crowd advisor over the single advisor. Thus, while the initial hypothesis that the crowd advisor would be widely used over the single advisor was <u>not</u> supported, there are some individual preferences in use of advice from different types of advisors. This may have critical implications for the design of recommendation systems that will be used by specific individuals.

Trust and Perceptions of Accuracy for Each Advisor Type

Beyond predicting differential use of advice across advisor types, *The Social Consensus* hypothesis also predicts differences in perceptions of accuracy and feelings of trust. To assess these factors, participants were asked two questions regarding their interactions with the advisors. First, participants were asked to estimate the advisors' accuracy. A key source of decision-making error is misestimating the abilities of those around us (e.g., Kruger & Dunning, 1999; Dunning, 2011). Thus, participants' ability to accurately assess an advisors' accuracy is hypothesized to be a key predictor of their subsequent advice use.

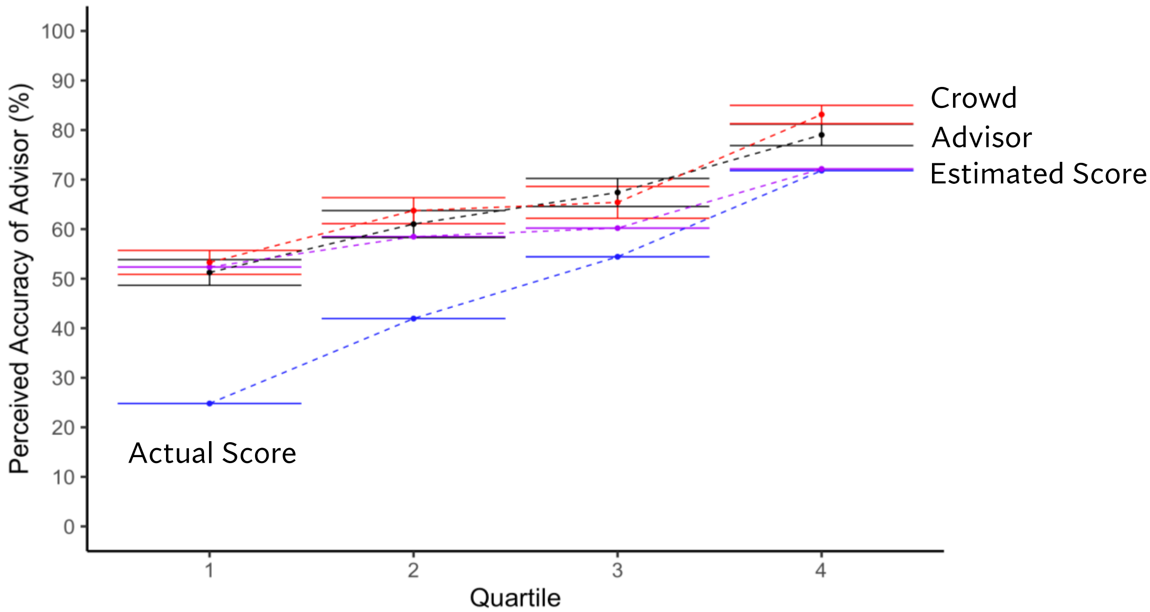
Second, participants were asked to estimate their trust in the advisor. Trust is defined as participants' willingness to believe that the information that an advisor provides is accurate to the best of the advisor's ability and is given with positive intentions (Hodges, 2014). Thus, participants may perceive an advisor as accurate, but not necessarily trust the advisors' intentions. Consequently, they may not trust nor use the advisor's advice.

Participants' ratings of their trust in the crowd's answers (M = 3.31, SD = 1.14 on a 1 (not willing at all) to 5 (completely willing) Likert scale) were significantly higher than that of their trust in single advisor's answers (M = 3.15, SD = 1.13 on a 1 to 5 Likert scale), $t(223)$ = 2.59, p = .010, d = .14. There was also a slightly higher rating of accuracy for the crowd (M = 65.54%, SD = 21.89%) than the single advisor (M = 63.75%, SD = 21.88%), but this was <u>not</u> a statistically significant difference, $t(220)$ = 1.44, p = .151, d = .08.

Per the Dual Burden account (Dunning, 2011), participants' ability to assess the performance of others is hypothesized to be affected by their domain skill (i.e., task score), with individuals with higher skill having a better ability to assess the performance of others. To test this hypothesis, differences in participants' ability to assess the advisor's accuracy due to their score was examined. The results are illustrated in Figure 8.

Figure 8.

Examination of Low versus High Scoring Participants' Ability to Assess Advisor's Accuracy



Note. The red line represents participants' estimates of the crowd advisor's accuracy. The black line represents the participants' estimates of the single advisor's accuracy. The purple line represents participants' estimates of their own accuracy (as captured at the end of the task). The blue line represents their actual score on the task (based on their pre-advice accuracy).

One-way Analyses of Variances (ANOVAs) were used to assess the effect of participants' score quartile on their ability to estimate the accuracy of the crowd and the single advisor. A Levene's test to assess for violations to the assumption of homogeneity of variance indicated that the crowd advisor ANOVA violated the assumption of equal variances, $F(3, 220) = 3.39$, $p = .019$, so an ANOVA with Welch's correction was used. The Levene's test for the single advisor ANOVA indicated no violations, $F(3,220) = 1.45$, $p = .229$, so a standard one-way ANOVA was used.

A clear effect of participants' task score can be seen for both the crowd advisor, $F(3, 118.96) = 34.91$, $p < .001$, and the single advisor, $F(3, 220) = 20.11$, $p < .001$. Those in the bottom 25% of participants estimated the crowd advisor as being only 53.3% accurate on average (*SD* = 19.4%), while the single advisor was rated as 51.3% accurate on average (*SD* = 20.7%). The top 25% of participants, on the other hand, estimated the crowd advisor as being 83.1% accurate (*SD* = 13.2%) and the single advisor as being 79.02% accurate (*SD* = 15.4%).

Due to the violation of the homogeneity of variance assumption for the crowd advisor ANOVA, a Games-Howell post-hoc comparison was used to assess differences in estimates between the top 25% and bottom 25% of participants. This analysis indicated that the bottom 25% of participants had 29.9%, $p < .001$ lower estimates of accuracy for the crowd advisor than the top 25% of participants. As the single advisor ANOVA indicated no violation to the assumption of homogeneity of variance, a Tukey post-hoc comparison was used instead of the Games-Howell technique. For estimates of the single advisor's accuracy, the bottom 25% of participants estimates were 27.8%, $p < .001$ lower on average than the top 25% of participants.

Chapter V Discussion

This chapter examined the *Social Consensus hypothesis*, which states that, due to the normative and informational social influence a crowd exerts, individuals are more likely to believe in the crowd advisor and use their advice, compared to a lone advisor. Three predictions were associated with the Social Consensus hypothesis. Participants were expected to trust the crowd advisor more, perceive the crowd advisor as more accurate, and ultimately use more advice from the crowd advisor. The results of these predictions will be examined in detail below.

Prediction 1: Greater Advice Taking from the Crowd's Advisor

Participants did <u>not</u> use significantly more advice from the crowd than from the single advisor. The probability of taking advice in this study for the single and crowd advisors ranged from approximately **24.8%**(95% CI [.209 – .290]) to **25.7%** (95% CI [.220 – .299]) respectively, which is consistent with the low rates of 20% to 30% seen in other studies (e.g., Harry & Fischer, 1997; Soll & Larrick, 1999; Soll & Larrick, 2009; Duan, Gu, & Sun, 2016; Pescetelli, Hauperich, & Yeung, 2021). The large sample size and the small 95% confidence intervals highlight that the data were precise enough to detect a meaningful increase in advice taking with a crowd advisor. Thus, regardless of who is providing the advice, advice taking is likely to remain low on average.

However, individual differences exist in participants' preference for one advisor over the other. While approximately 40% of participants displayed no preference for one advisor over the other, approximately 28% of participants preferred to use the single advisor, while approximately 31% of participants preferred the crowd advisor. Therefore, designers of recommender systems should assess users' willingness to use one advisor over the other to best utilize individual preference to enhance advice use.

Predictions 2 & 3: Trust and Perceptions of Accuracy

Of the three predictions associated with the Social Consensus hypothesis, only the prediction that an individual would trust the crowd advisor more than the single advisor was supported by empirical results. However, the difference in ratings was quite small (crowd: M = 3.31; advisor: M = 3.14) and the average rating for both advisors was near midpoint of the 5-point Likert scale.

The final prediction stated that individuals will rate the crowd advisor as being more accurate than the single advisor. Participants rated the crowd advisor as slightly more accurate than the single advisor, albeit non-significantly. Therefore, the take-away message of this section is that while the crowd is seen as slightly more trustworthy and accurate, the large differences predicted by the Social Consensus hypothesis were not seen. Therefore, leveraging a crowd-based advisor to enhance advice taking may not elicit desired results.

Participants' Task Score and Advice Taking

It is of interest to note that, despite both advisors providing 100% correct advice, participants on average rated both advisors as having accuracies in the mid-sixty percent range (approximately 64% to 66%). An important qualifier of participants' estimates of the advisors' accuracy is individual's domain ability (i.e., their task score). Participants who were in the top 25% of participants rated the single and crowd advisors as being approximately 79% to 83% accurate, respectively. This is contrasted by the estimates of the bottom 25% of participants, who rated the single advisor as being approximately 51% accurate and the crowd advisor as being approximately 53% accurate. This is a 28% to 30% difference in estimates between the highest and lowest skilled.

At first glance, it appears that one's higher domain ability seemingly endows these individuals with a greater ability to assess the accuracy of others, which allows them to better assess the veracity of the decision-making of those around them. However, a new pattern emerged when participants' self-estimates of performance (based on their post-task judgments) were plotted against participants' estimates of the advisor's accuracy. Participants placed the advisor's accuracy close to their own, which suggests that they are using a heuristic judgment to estimate the advisor's accuracy based largely on the percentage of agreement between the advisors' answers and their own. When the advisor disagrees with the participant, rather than the participant using that disagreement as a cue to rethink their answer, they are treating this inconsistency as the advisor's error. This was seen even in the high-skilled, who were expected to make more accurate judgments about the abilities of others. The role skill plays in advice taking patterns will be further explored in *Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results*.

Chapter V: Conclusion

Despite 100% accuracy of the advice, participants resisted taking advice on their error trials. Participants had an approximately 26% probability of accepting advice on average, which is aligned with previous research highlighting low levels of advice utilization. However, there was significant individual variation in the probability of accepting advice, ranging from 10% to 80% probability of accepting advice. This is also in-line with previous research demonstrating individual variation ranging from 0% to 72% (Pescetelli, Hauperich, & Yeung).

While the crowd manipulation was hypothesized to enhance advice taking, per the Social Consensus hypothesis, the crowd's advice was not used significantly more than that of the single advisor, with advice utilization for both advisor types remaining extremely low. In addition, while participants with high scores on the task were approximately 30% more accurate in assessing the advisors' capabilities than those with low task scores, they still placed the advisor's accuracy near themselves and not at the advisor's true accuracy of 100%. This is largely because disagreements with the advisor are being attributed to advisor error, which leads even the high-skilled to rate the advisors' accuracies below their true level.

Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results

The goal of this chapter is to explore differences in advice taking based on participants domain skill and metacognitive ability. Prior advice taking research has largely ignored the role of skill and metacognitive ability in advice taking, despite other literatures finding large overconfidence effects in individuals' ability to accurately assess their capabilities (e.g., Kruger & Dunning, 1999). Overestimating one's ability is likely to hinder individuals' advice use, as they will not be able to accurately identify when they are likely to be incorrect and in need of help (i.e., poor metacognitive sensitivity). Skill is expected to directly impact participants' metacognitive sensitivity, as skill allows individuals to accurately assess differences between their correct and incorrect trials (i.e., the Dual Burden hypothesis; Dunning, 2011).

Two hypotheses are considered in this chapter: the *High Confidence Leads to Advice Blindness* hypothesis and the *Skill-Based Advice Taking* hypothesis. The High Confidence Leads to Advice Blindness hypothesis is based on prior research by Sanchez, Benson, and Ruthruff (2023), which has highlighted that regardless of participants domain skill, when in a high confidence state, all individuals are blind to their errors (i.e., Universal Error Blindness hypothesis). The High Confidence Leads to Advice Blindness hypothesis, which will be referred to as the *High Confidence* hypothesis for short, states that people are especially sensitive to their internal subjective feelings of confidence in their initial answer, and relatively insensitive to external cues, such as disagreement between their and the advisor(s)' answer. Thus, they are unlikely to accept advice when in a high-confidence state, as this confidence blinds them to the need for advice. The strong version of this High Confidence hypothesis assumes that this advice blindness is a general feature of human cognition, applying to people of all skill levels.

The *Skill-Based Advice Taking* is based on the Dual Burden account by Dunning (2011). The Skill-Based Advice Taking hypothesis states that, due to their superior ability to identify when they are incorrect (i.e., metacognitive sensitivity), the high-skilled are better able to use advice as a cue to rethink their initial answer. Thus, even when they are initially in a high-confidence state, they are more likely to use advice than are those with low skill. Throughout the rest of this section, the *Skill-Based Advice Taking* hypothesis will be referred to simply as the "Skill-Based hypothesis."

Participants' Ability to Accurately Assess Their Performance

To determine how participants' ability to accurately estimate their performance varies across skill levels, they were first sorted and divided into score[3] quartiles, per Kruger and Dunning (1999). Thus, Quartile 1 represents the bottom 25% of participants, while Quartile 4 represents the top 25% of participants. After this division, the average estimated performance was plotted against the average achieved performance for each quartile, following Kruger and Dunning (1999) convention. Although the use of quartiles has sometimes been criticized on the grounds that quartile-based analyses are less powerful than analyses treating skill as a continuous variable, it has the important advantage that it makes it easy to visualize data patterns.

Two outcome measures were used to examine participants' ability to accurately assess their performance. The first is percentile estimate overconfidence. Percentile estimates were captured by asking participants out of 100 other participants completing

[3] Score was calculated based on the accuracy of the initial answer participants had given when answering the logic tasks questions.

this study, how many individuals they believed they would score *higher* than. To calculate their overconfidence, their achieved percentile was subtracted from their estimated percentile. The second outcome measure is participants' percentage correct overconfidence. Percentage correct estimates were captured by asking participants what percentage of the logic task they believe they would answer correctly. Overconfidence was calculated by subtracting participants' achieved percentage correct from their estimated percentage correct. Table 4 contains the means and standard deviations for these outcome measures for each performance quartile. The mean number of errors perpetrated by each quartile can be found in Table 4.

As the Dunning-Kruger effect the greatest differences in accurately estimating one's performance thought to lie in the behaviors of the lowest skilled (the bottom quartile) and the highest skilled (the top quartile) individuals, analyses will highlight the difference between these individuals in their ability to estimate their performance and identify their errors.

Table 4.

Mean and Standard Deviations for Outcome Measures by Performance Quartile

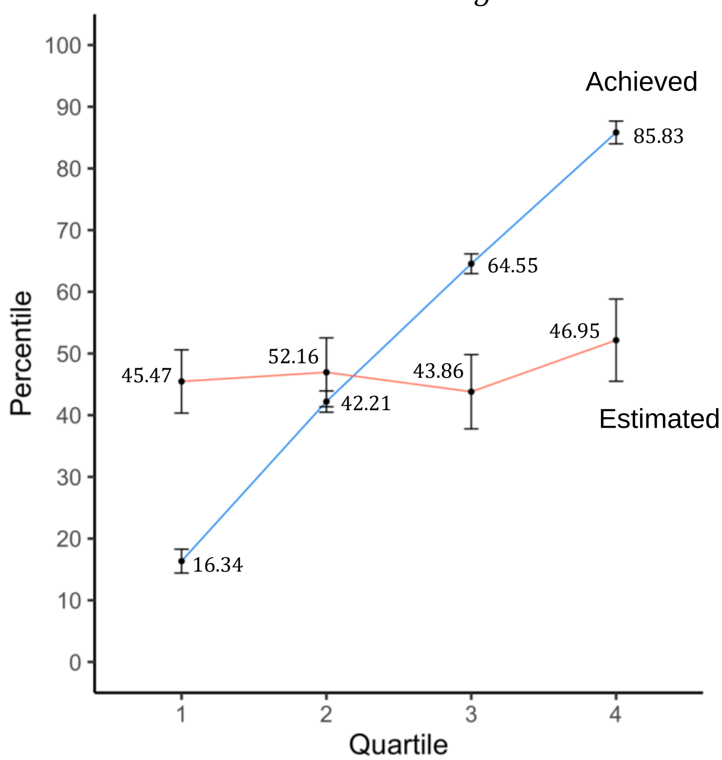
Score Quartile (Task Score Range)	N	Mean Task Score	Mean Percentile Est. Overconfidence	Mean % Correct Est. Overconfidence
Q1 (8.3% - 33.3%)	64	24.80 % (6.57%)	29.13 (23.05)	27.54% (22.71%)
Q2 (33.4% - 45.8%)	59	41.95% (3.53%)	4.74 (23.59)	16.53% (19.48%)
Q3 (45.9% - 58.3%)	50	54.42% (3.30%)	- 20.75 (22.58)	5.78% (20.51%)
Q4 (58.4% - 87.5%)	51	71.81% (7.20%)	- 33.67 (23.53)	.34% (13.64%)

Note. Negative values indicate underconfidence, while positive values indicate overconfidence for mean percentile estimation and mean percentage correct overconfidence columns. Values in parentheses are standard deviations.

As illustrated in Figure 9, participants' percentile ranking estimates compared to their achieved percentile matches the typical Dunning-Kruger pattern of overconfidence in Quartile 1 and underestimation in Quartile 4. Participants in the bottom quartile *overestimated* their overall percentile ranking ($M_{Estimated}$ = 45.47th) when compared to their actual ranking ($M_{Achieved}$ =16.34th), paired $t(63) = 10.11$, $p < .001$, $d = 1.26$. On the other hand, individuals in the top quartile *underestimated* their percentile ranking ($M_{Estimated}$ = 46.95th) compared to their achieved percentile ($M_{Achieved}$ = 85.83th), paired $t(50) = - 10.22$, $p < .001$, $d = - 1.43$. Thus the standard Dunning-Kruger effect has been replicated in this sample (e.g., Kruger & Dunning, 1999; Ehrlinger et al., 2008). On the surface at least, this finding appears to support the assumption of the skill-based model that the high-skilled are better able to judge their performance.

Figure 9.

Estimated Percentile Rank Plotted Against Achieved Percentile Rank

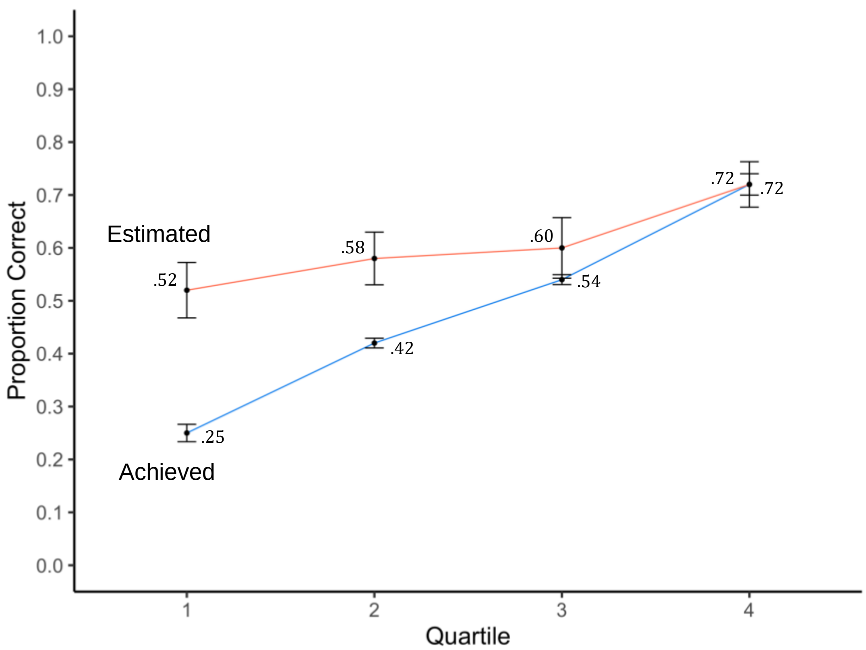


Note. This plot represents percentile ranking estimates, as measured by asking participants "out of 100 other psychology participants completing this task, how many do you believe you would score higher than?" after the task. Error bars represent 95% confidence intervals.

Figure 10 highlights participants' estimates of their percentage correct on the logic task in comparison to their achieved percentage correct. As expected, those in the bottom quartile still *overestimated* their performance, as they estimated they would answer 27.54% more of the task correctly than they actually did, paired $t(63)= 9.70$, $p < .001$, $d = 1.21$. However, the top quartile accurately predicted that would answer approximately 72% of the task correctly, paired $t(61) = .18$, $p = .858$, $d = .03$. These results are similar to the original Dunning-Kruger pattern observed for estimates of task score (Kruger & Dunning, 1999), though they asked participants the number of questions they would answer correctly rather than proportion of the task.

Figure 10.

Estimated Logic Task Score (%) Against Achieved Score (%)



Note. This plot represents the global accuracy estimates of performance, in which participants were asked to estimate the percentage of the task they had answered correctly at the end of the task. These estimates were then plotted against participants' accuracy, which was their score dived by total number of questions. Error bars represent 95% confidence intervals.

Are People Aware of Their Advice Use?

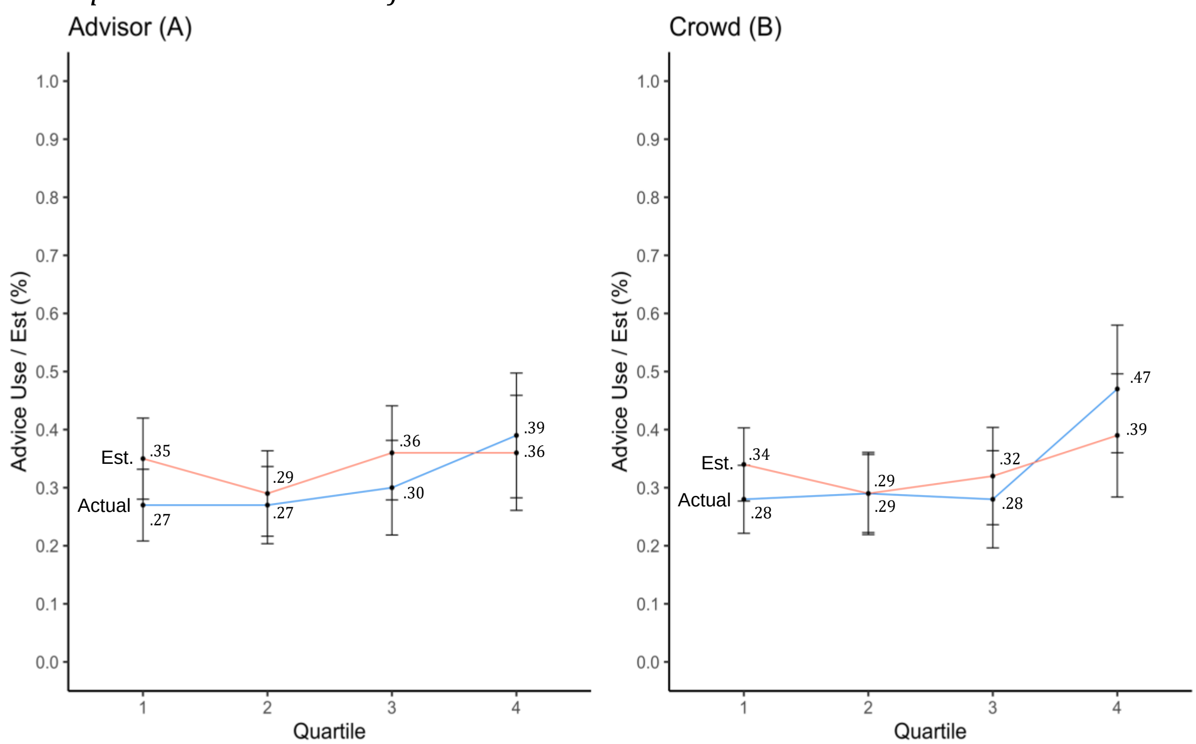
Next, the empirical question of whether participants are aware of their actual advice use or if they misestimate their use was examined, as highlighted in Figure 11. To examine individuals' accuracy in estimating their advice use, a multiple linear regression was constructed. Participants' task score and the type of advisor (single or crowd) were entered as predictors into the model. The dependent variable was calculated using the following equation:

$$Advice\ Taking\ Overconfidence\ =$$

$$Estimated\ Percentage\ of\ Advice\ Used\ -(\frac{Number\ of\ Times\ Advice\ Taken}{Total\ Number\ of\ Errors})$$

Advice taking overconfidence scores ranged from -1 to 1, with a difference score of -1 indicating that the use of advice is greater than the *estimated use* of advice, while a difference score of 1 indicates that the *estimated use* of advice is greater than actual use. As the type of advisor is a categorical variable, the single advisor was coded as the reference group.

Figure 11.
Participants' Estimated Use of Advice Versus Their Actual Advice Use



Linear regression assumptions were assessed, including the presence of non-linearity, heteroscedasticity, the presence of influential values or outliers, and non-normality of residuals. The results of a Normality of Residuals QQ-plot indicated that the residuals violated the assumption of normality. However, the assumption of normality can be relaxed when N is 50 or larger due to the Central Limit Theorem (Pek, Wong, & Wong, 2018). The Central Limit Theorem states that the sampling distribution of estimates will converge towards a normal distribution the larger the sample size is, making a linear model robust to the violation of normality without transforming the data (Pek, Wong, & Wong, 2018). No other violations were found.

As highlighted in Figure 11, those who scored higher on the task were more likely to *underestimate* their advice use, while those who scored lower on the task were more likely to *overestimate* how much advice they had used, regardless of the advisor type. Results of a multiple regression analysis used to analyze over and underestimation in advice taking indicated a collective significant effect of task score and advisor type on advice taking overconfidence scores, $F(2, 445) = 8.51$, $p < .001$, $R^2 = .04$. Examining individual predictors further, only participants' task score was a significant predictor of advice taking overconfidence scores ($\beta = -0.18$, $t(445) = -3.94$, $p < .001$). No significant effect of advisor type was found ($\beta = -0.06$, $t(445) = -1.21$, $p = .225$).

Testing the High Confidence Advice Blindness & Skill-Based Hypotheses

This section starts by testing the assumptions of the High Confidence and Skill-Based Hypotheses. The first assumption tested is the High Confidence assumption that individuals of all skill levels display high pre-advice confidence on their error trials due to universal error blindness. It is hypothesized that this assumed high confidence on one's error trials then prevents individuals from seeing the need for advice and using it to correct their error(s).

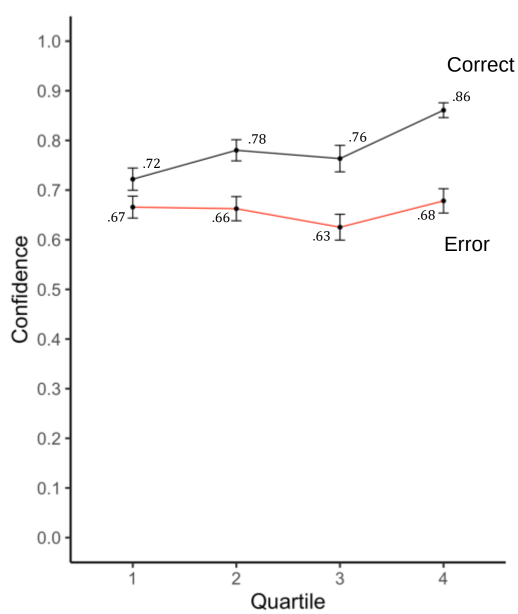
The second assumption tested in this section is the assumption of the Skill-Based hypothesis that individuals with higher skill have better metacognitive sensitivity. Metacognitive sensitivity is the ability to distinguish between one's correct and incorrect trials (Fleming & Lau, 2014). Because they have a better sense when their thinking might be in error, they can take the presented advice as a nudge to rethink their initial answer.

Assumption 1: Assessing the Degree of High Confidence Errors Across Score Quartiles

The work of Sanchez, Benson, and Ruthruff (2023) has highlighted that, regardless of an individuals' skill level, they are often highly confident on their error trials (the Universal Error Blindness hypothesis). This effect runs counter to the Dual Burden hypothesis (Dunning, 2011), which suggests that this blindness to one's errors should be found mainly in lower scoring individuals. Figure 12 highlights each score quartile's average confidence on trials for which they were correct and for which they had made an error.

Figure 12.

Average Confidence on Error and Correct Trials by Performance Quartile



Confidence on Error Trials. Levene's test for the error trials demonstrated that the assumption of equal variance was not violated ($F(3, 220) = 0.14$, $p = .938$), so a standard one-way ANOVA was used to examine the effect of score quartile on confidence on error trials. The one-way ANOVA revealed that there was <u>not</u> a significant effect of score quartile on error trial confidence, $F(3, 220) = 0.811$, $p = .489$, = .01. A Tukey post-hoc test revealed a - .01 difference in error confidence between the bottom and top quartile participants, 95% CI [-.10, .08], $p = 0.982$. Thus, there is no evidence that the bottom quartile was any more confident than the top quartile while making an error. Instead, the results fit the Universal Error Blindness hypothesis in this sample.

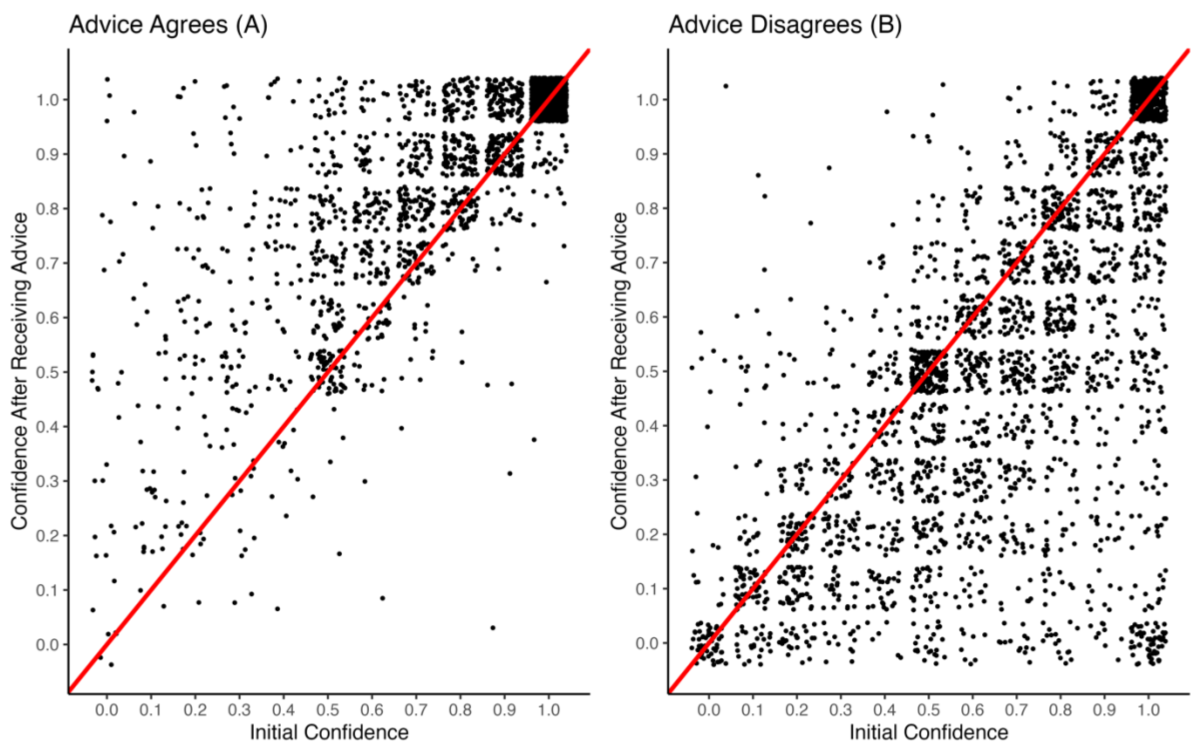
Confidence on Correct Trials*.* The main analysis of interest in this investigation is the difference between quartiles regarding their confidence on error trials. However, confidence on correct trials is also reported for completeness. Levene's test indicated significant differences in variance across quartiles, $F(3, 220) = 4.37$, $p = .005$, for confidence on correct logic trials. A one-way ANOVA with Welch's correction was used. Results indicated that score quartile was a significant predictor of confidence on correct logic task trials, $F(3, 118.93) = 10.39$, $p < .001$, and it explained 21% of the variance in correct trial confidence estimates, $\eta^2 = .21$.

A Games-Howell post-hoc analysis indicated that confidence on logic correct trials was significantly higher in top quartile participants (Q4) than in bottom quartile participants (Q1) by an average of 13.89%, $p < 0.001$. Participants in Q4 were also significantly more confident on their correct trials than their peers in the second quartile (Q2), with a mean difference of 8.07%, $p = .013$. This trend persisted with third quartile (Q3) participants, with Q4 participants an average of 9.75% more confident on correct trials, $p = .011$. There were no significant differences between other quartiles for confidence on correct trials. Q4's higher confidence on their correct trials may be due to greater metacognitive ability (i.e., the ability to know when is correct versus incorrect). This will be explored in-depth in the next section.

Exploring Changes in Confidence Pre-Advice and Post-Advice. Before testing the competing hypotheses, the role of confidence in advice acceptance was examined. Confidence was captured when participants first gave their answer (pre-advice confidence) and after receiving the advice (post-advice confidence). Figure 13 depicts changes in confidence in one's original answer prior to seeing the advice (x-axis) and after seeing the advice (y-axis).

Figure 13.

Individuals' Confidence in Their Original Answer Pre- and Post-Advice



When the advice agreed with the participants' initial answer, confidence increased (or remained high if it was high to begin with). The opposite pattern was seen when the advice disagreed with the participants' initial answer. In these cases, participants demonstrated a decrease in confidence from their pre-advice confidence level as a result of seeing the advice.

Figure 14 highlights participants' changes in confidence when they decided to change their answer versus when they did not on their error trials. Panel A contains a scatterplot of participants' confidence prior to receiving advice (i.e., pre-advice confidence) plotted against their confidence in their initial answer after receiving the advice (i.e., post-advice confidence) when they accepted advice. Panel B is a histogram of participants' post-advice confidence in their initial answer when they accepted advice. Panel C is similar in structure to Panel A but plots pre-advice and post-advice confidence when participants did not accept the advice. Panel D similarly plots participants' post-advice confidence in their initial answer when they did not accept advice.

Figure 14.

The Impact of Participants' Confidence After Receiving Advice on Answer Changing

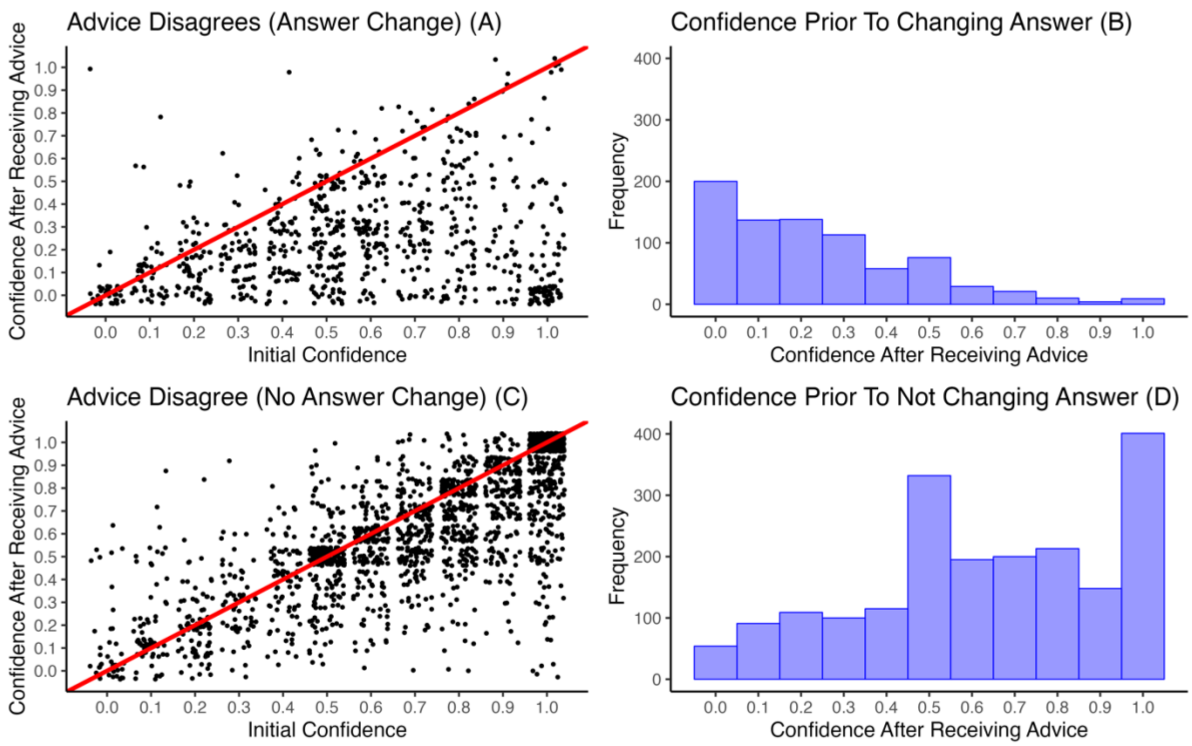


Figure 14 highlights when post-advice confidence was low (0% to 30%), participants were found to take advice more frequently than when then when post-advice confidence was higher (70% to 100%). Thus, there is a tight connection between these variables, as would be expected. Participants remaining confident in their original answer after seeing conflicting advice have clearly not taken the advice.

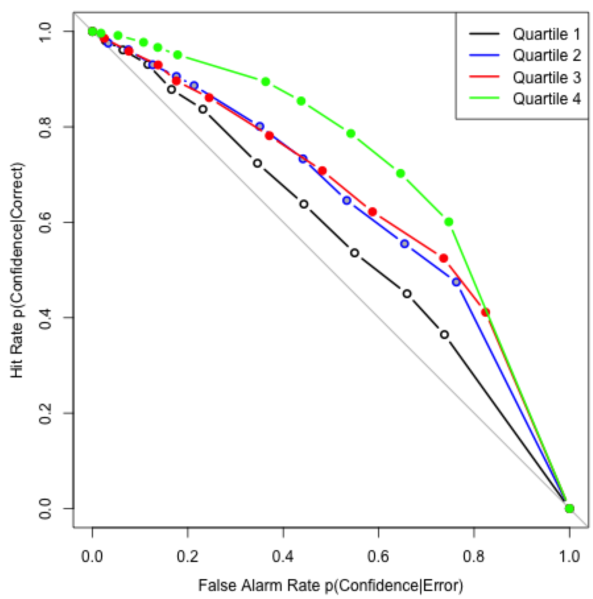
Assumption 2: The High-Skilled Have Superior Metacognition

Participant's metacognitive ability was examined using a Type II receiver operating curve (ROC). A ROC curve plots an individual's *hit rate* (proportion of times a given confidence level was used on a correct trial) against their *false alarm rate* (proportion of times a given confidence level was used on an error trial; Fleming & Lau, 2014). ROC curves are thought to be "bias-free," in that they are not biased by differing degrees of caution in using specific confidence ratings (criterion; Fleming & Lau, 2014).

As low task score (i.e., low skill) is hypothesized to lead to lower metacognitive ability, ROC curves were calculated for each score quartile to ascertain whether those in the bottom quartile demonstrated poorer metacognitive sensitivity then their top quartile counterparts. The ROC curves for each quartile can be found in Figure 15.

Figure 15. ROC Curves for Each Score Quartile

ROC Curves for Each Score Quartile



The area under the ROC curve (AUC) is a measure of an individuals' metacognitive sensitivity (Fleming & Lau, 2014). The higher the AUC value, the more metacognitive sensitivity a person is thought to have. An AUC of .5 is considered at chance (i.e., no discrimination between one's correct and error trials), while an AUC of 1 indicates perfect discrimination between one's correct and error trials. As demonstrated in Figure 15, the AUC for the bottom quartile (Q1) is lower than that of the top quartile (Q4), indicating lower metacognitive sensitivity within this group. The mean AUC for those in the bottom quartile was .64 (*SD* = .12), while those in the top quartile had a mean AUC of

.73 (SD = .12). Quartiles 2 and 3 had a mean AUC of .65 (SD = .10) and .69 (SD = .12), respectively.

A further examination of the relationship between task skill and metacognition was conducted using participants' task score (a continuous variable) rather than their score quartile (a nominal variable). This offers a series of benefits including greater statistical power due to increased precision of the variable, as well a simpler, more informative interpretation of the results (Lazic, 2008).

Data met the necessary assumptions for a linear regression (linearity, homogeneity of variance, normality of residuals, and outliers or influential points) and no violations were identified. Logic score was mean centered before being inputted into the model. The model was significant, with score explaining 7.05% of the differences in AUC values $F(1, 222) = 16.85$, $p < .001$, $R^2 = .0705$. Results indicated a positive relationship between score and AUC, with a one unit increase in score leading to a .18 increase in AUC, $b = .18$, $t(222) = 4.11$, $p < .001$.

However, there is an important limitation to be highlighted when comparing metacognitive ability between high and low-skilled individuals, as highlighted in Sanchez (2021). Current statistical methods for examining metacognitive sensitivity cannot sufficiently separate task performance from calculations of participants' metacognitive ability (Fleming & Lau, 2014; Fleming, 2017; Vuorre & Metcalf, 2021). This bias is driven by participants' d-prime and criterion values (Fleming & Lau, 2014; Fleming, 2017).

D-prime is the standardized difference between the mean of the distribution of participants' confidence on correct trials and the mean of the distribution of confidence on error trials. Criterion is participants' willingness to use the high versus low-end of the confidence scale when they are correct. Participants with higher task skill appear to have better metacognitive ability solely by the fact that they are correct more often, which makes it appear that they have higher separation between their distribution of correct and error confidences (Galvin, 2003; Fleming & Lau, 2014; Fleming, 2017).

Intuitively, this makes sense. Imagine two individuals, Individual A and Individual B, who both state that they are 100% confident on 10 questions. However, Individual A scores 8 out of 10 and individual B scores 6 out of 10. While both participants committed high confidence errors on their incorrect trials, Individual B appears to have greater blindness to their errors due to the fact that their error rate is double (40% of trials) that of Individual A (20% of trials). Consequently, Individual A appears to have better calibration in regard to their accuracy, because the 80% of the time they said they were correct, they actually were.

However, better methods for calculating metacognitive sensitivity, independent from skill, does not yet exist, unless the task at hand is a 2-answer forced-choice task (Vuorre & Metcalf, 2021). ROC curves, which were used in this analysis, are recommended for controlling for the bias introduced by the criterion (i.e., the overall propensity for using the high or low-end of the confidence scale), but do not sufficiently control for the bias introduced by task skill. Therefore, this work could not provide a perfect test of the relationship between task skill and AUC, with some inflation likely to exist in the high-skilled's calculated AUC. This is a limitation both of this work and of the field at large.

The Battle Between Hypotheses: Resolving the Skill-Based & High Confidence Hypotheses Debate

The key analysis of interest in this chapter is determining if the Skill-Based or High Confidence hypotheses better explain advice taking patterns seen in this sample. The

High Confidence hypothesis states that when an individual is in a high confidence state, advice will be ignored, regardless of participants' skill level, due to a blindness that the confidence induces to the need to accept feedback. Whereas the Skill-Based hypothesis states that ignoring advice when in a high confidence state should only occur in low-skilled individuals. To test these competing hypotheses, assumptions for each hypothesis first had to be assessed before a direct comparison of hypotheses could be carried out.

First, for the High Confidence hypothesis to be viable, individuals must commit high confidence errors on their incorrect trials. This assumption was supported, with both the low and high-skilled (as measured by task score) committing high confidence errors. Confidence on error trials ranged from 63% to 68%, with no significant difference between skill groups in their average error confidence.

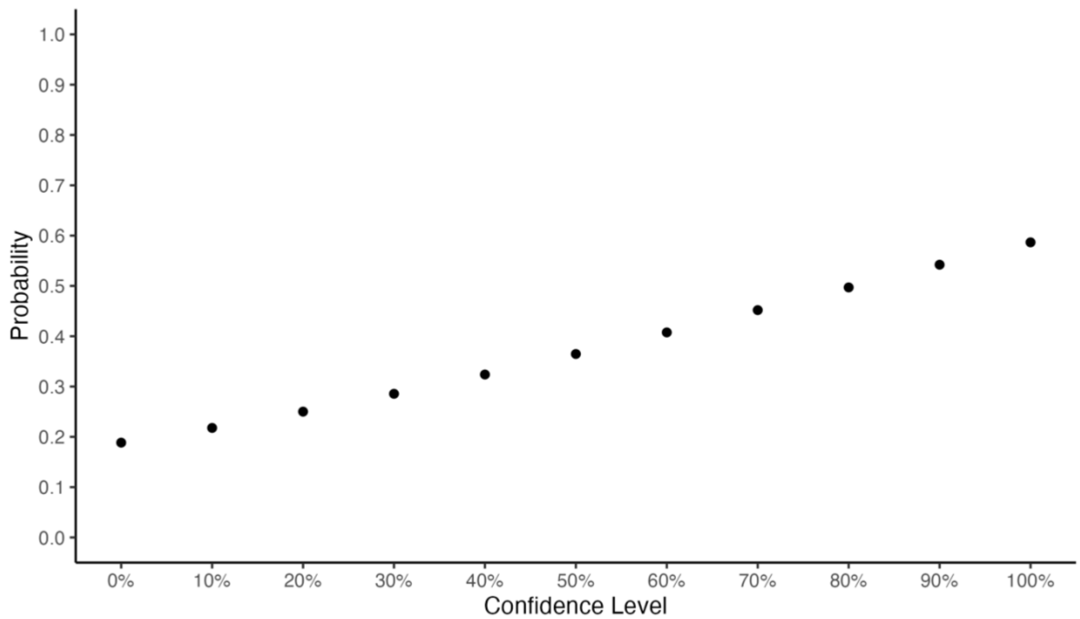
For the Skill-Based hypothesis to be viable, the assumption that participants with higher skill must be better able to detect their errors (i.e., have better metacognitive sensitivity) must be met. This assumption was supported, with those in the top 25% of participants (i.e., Quartile 4) having the greatest metacognitive ability of all the score quartiles. Subsequent analyses examining advice taking in highly skilled and high metacognitively sensitive individuals found that greater skill and AUC predicted advice use. Thus, assumptions of both the High Confidence and Skill-Based hypotheses have been empirically supported and a direct comparison of these hypotheses is conducted in the next section.

Does Being in a High Confidence State Predict Lower Advice Taking? Testing The High-Confidence Hypothesis

While the primary analysis of interest centers on participants' error confidence and its predictability of accepting advice, it's an important step to first explore the diagnosticity of participants' confidence on their probability of being accurate. A logistic regression was run using pre-advice confidence (uncentered) and regressing it on accuracy (0 = incorrect, 1 = correct). The result of this analysis is highlighted in Figure 16.

Figure 16.

The Probability That a Participant was Correct at Each Level of Confidence



As highlighted in Figure 16, as participants' confidence level increases, so does the probability of them being correct, $b = 1.81$, $se = .11$, $p < .001$. Participants were least likely to be accurate when their confidence was at 0% (18.9% probability of being accurate) and had the greatest likelihood of being accurate at the 100% confidence level (58.7% probability of being accurate). Next, the role confidence plays in participants' likelihood of accepting advice will be explored.

Model Construction and Fit. To test the High Confidence hypothesis, a multi-level logistic regression model was constructed to analyze whether being in a high confidence state prior to receiving advice leads to lower advice taking. Accepting advice on one's error trials (0 = did not accept advice, 1 = accepted advice) was the outcome measure of interest, with random intercepts for each participant included in the model in conjunction with random slopes based on participants' grand-mean centered pre-advice confidence ratings. The following equation describes the model:

$$\boldsymbol{p(Switching\ to\ Advisor's\ Answer)_{ij}} = \beta_{0j} + \beta_{1j} Pre - Advice\ Confidence\ Rating_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

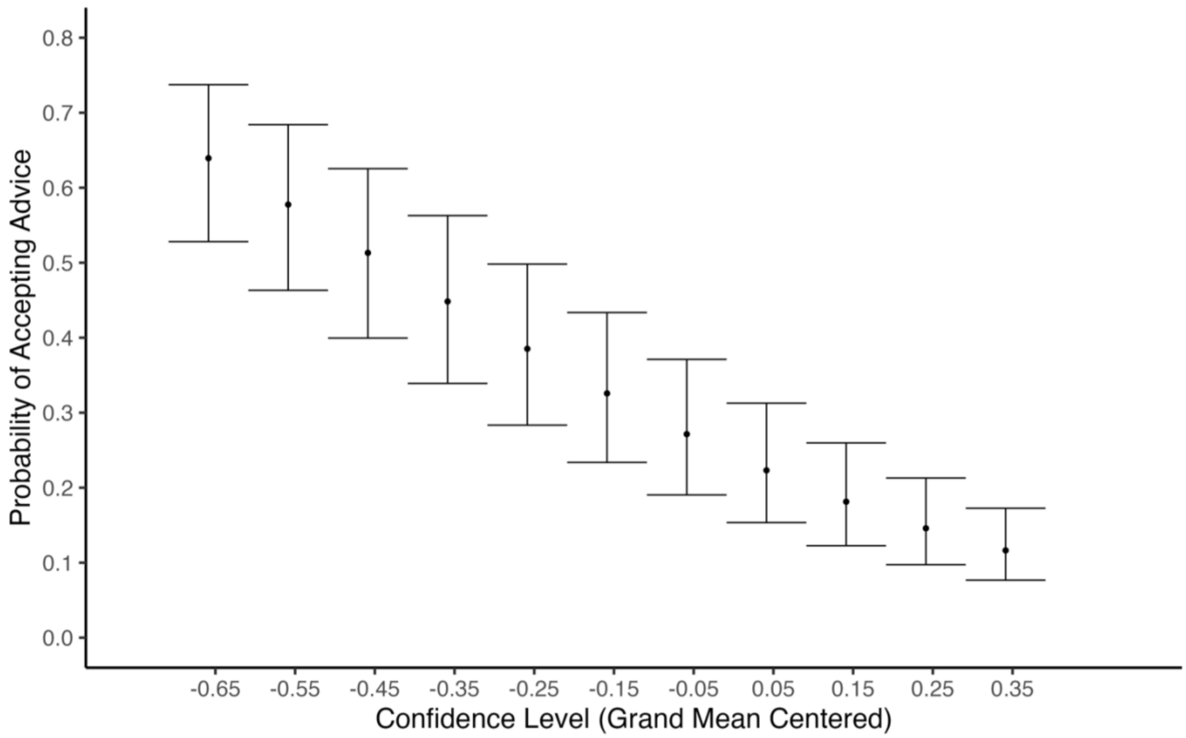
where *i* represents each logic question nested within participant *j.* β_{0j} represents the random intercept for each participant, while β_{1j} represents the random slope of pre-advice confidence for each participant *j*. 224 participants were included in the model with 2753 trials. The model was fit using maximum likelihood (Laplace Approximation) using the lme4 package's glmer() function (Bates et al., 2015).

Does Higher Pre-Advice Confidence Leads to Lower Advice Taking? The fixed intercept indicates that when confidence is at the grand mean level (66% confident), advice taking will be low, β_0 = - 1.14 log-odds, *se* = .10, $p < .001$. Converting to probabilities, individuals had a 24.2%, 95% CI [0.21 – 0.28] probability of accepting advice when their advice was at the mean level. However, individuals had significant variation in their probability of accepting advice, with random intercepts having a standard deviation of 1.19 log-odds. The random intercept for each individual is plotted in Figure 18.

As predicted, the fixed effect of confidence was found to have a strong negative relationship with probability of accepting advice, β_1 = - 2.60, *se* = .23, $p < .001$, supporting the High Confidence hypothesis. The probability of an individual accepting advice at different confidence levels are illustrated in Figure 17.

Figure 17.

The Probability That Participants Accepted Advice at Differing Levels of Confidence

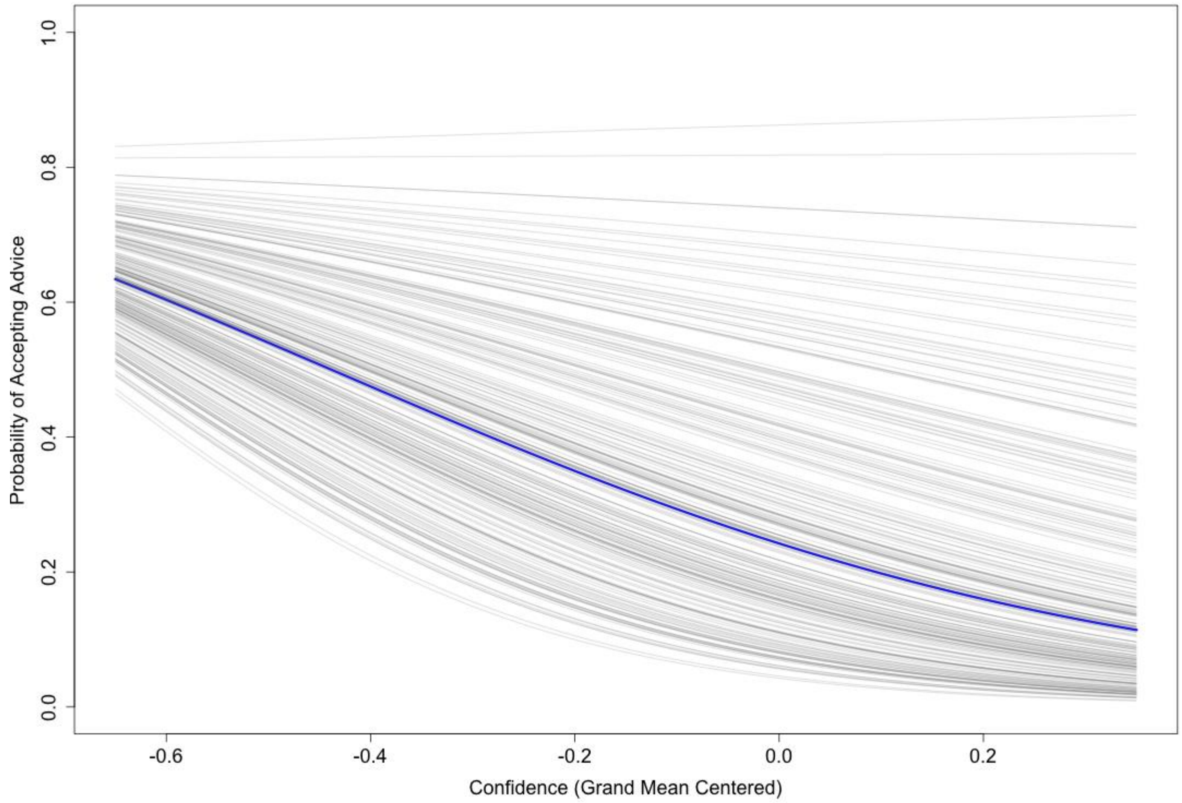


Note. Error bars indicate 95% confidence intervals based on the standard errors.

While in general, when in a high confidence state, individuals are less likely to accept advice, high individual differences exist. There is a standard deviation of .89 log-odds for individuals' random slopes from the fixed slope of - 2.60 log-odds for pre-advice confidence. The random slopes and intercepts for each participant were plotted to examine differences across participants in the probability of accepting advice when in a high versus low confidence state. This is illustrated in Figure 18. In general, when in a high confidence state, individuals are less likely to accept advice, but high individual variation exists

Figure 18.

Random Effects: Probability of Accepting Advice in a High Versus Low Confidence State



Do the High-Skilled Take More Advice in a High-Confidence State Than the Low-Skilled? Testing the Skill-Based Hypothesis

Model Construction and Fit. To test the skill-based hypothesis that the higher skilled are more likely to take advice when in a high confidence state than their low-skilled peers, due to their better ability to rethink their initial answers, a multi-level logistic regression model was used. Accepting advice (0 = did not accept advice, 1 = accepted advice) served outcome measure of interest, with fixed effects for participants' pre-advice confidence and logic score (both grand-mean centered). Random intercepts and random slopes (based on pre-advice confidence) were included for each participant. The following equation describes the model:

$$\boldsymbol{p(Switching\ to\ Advisor's\ Answer)_{ij}} = \beta_{0j} + \beta_{1j}Pre - Advice\ Confidence\ Rating_{ij} + \beta_2 Logic\ Score_j + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

where *i* represents each logic question nested within participant *j.* β_{0j} represents the random intercept for each participant, while β_{1j} represents the random slope of pre-advice confidence for each participant *j*. 224 participants were included in the model with 2753 trials. The model was fit using maximum likelihood (Laplace Approximation) using the lme4 package's glmer() function (Bates et al., 2015).

Odds ratios are commonly reported when examining logistic multi-level models, with odds ratio describing the impact of Y given a 1 unit change in x-variables. However, the confidence and score x-variables are on the .1-unit scale. Due to this discrepancy in units for the x-variables confidence and logic score, the 95% CI for these odds ratios have extremely large ranges. Therefore, log-odds are reported rather than the odds ratios, which are highlighted in Table 5. Due to this problem with odds ratios, the predicted probabilities will be used to examine the interaction between skill and confidence, which is central to answering if the high-skilled are better at taking advice when in a high confidence state.

Table 5.

Multi-level Model Results Testing the Skill-Based Model

Predictors	*Log-Odds*	*std. Error*	*p*
Intercept	-1.10	0.10	**<0.001**
Confidence[a]	-2.38	0.23	**<0.001**
Score[a]	1.23	0.53	**0.021**
Score x Confidence[a]	3.49	1.20	**0.004**
Random Effects			
τ_{00} (Variance of Random Intercepts)	1.36		
τ_{11} (*SD* of Random Slope)	0.62		
ρ_{01} (Correlation: Intercepts & Slopes)	-0.72		
N	224		
Number of Trials	2753		

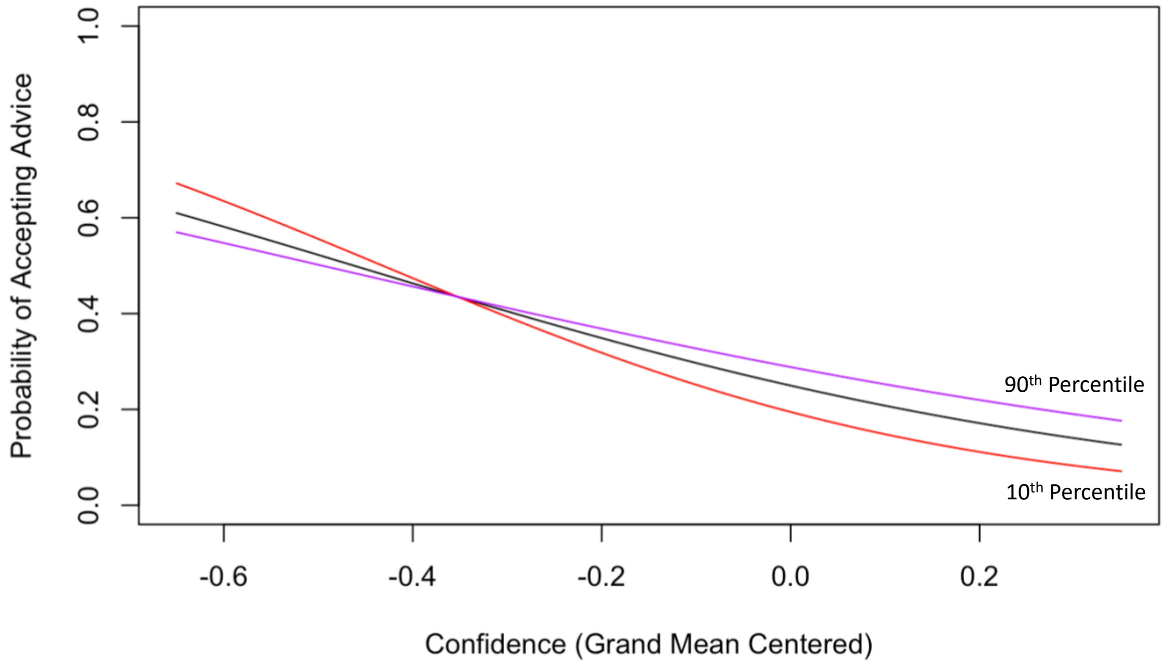
Note.
[a] Grand Mean Centered Variable(s)

Are the High Skilled Better at Accepting Advice in a High Confidence State?

To answer this question, the interaction between an individual's skill level and their confidence on the probability of accepting advice was examined. This analyze examined advice taking patterns for an individual in the 10th and 90th score percentiles. Participants in the 10th percentile represent an individual who scored approximately 21% correct (grand mean centered value = - .26) on the task, while those in 90th percentile represents an individual who scored approximately 71% correct on the task (grand mean centered value = .16). Values for pre-advice confidence and logic score for an individual in the 10th percentile and 90th percentile were calculated using the quantile() function in base R. The grand mean centered values were used to plot the results, which are highlighted in Figure 19.

Figure 19.

Predicted Probability of Accepting Advice from Individuals Who Scored in the 10th & 90th Percentiles



Note. The black line represents the fixed intercept and slope for the sample. The red line represents the intercept and slope of an individual who scored in the 10th percentile. The purple line represents the intercept and slope of an individual who scored in the 90th percentile.

As can be seen in Figure 19, when participants were in a low-confidence state, participants in the 10th percentile (red line) had a higher probability of taking advice than their 90th percentile peers (the purple line). Participants in the 10th percentile were 67.2% likely to accept advice when their confidence was low, whereas participants in the 90th percentile were 57.0% likely to accept advice in a low-confidence state.

However, as confidence increases above the mean level (66% confidence), the probability that either skill-level will accept advice decreases sharply. Participants in the 10th percentile were only 7.1% likely to accept advice when in a high confidence state, as opposed to individuals in the 90th percentile who were 17.6% likely to accept advice in a high confidence state. Per the interaction term included in the logistic regression multi-level model, these differences are statistically significant $b = 3.49$, $p < .001$.

Skill Helps Increase Advice Taking When in A High Confidence State, but is Metacognition the Source of this Increase? Investigating the Dual Burden Hypothesis

Research on the Dunning-Kruger effect suggests that skill and an individual's ability to know when they are likely to be incorrect (i.e., metacognitive sensitivity) should be key factors in promoting advice taking. High skill enhances individuals' metacognitive ability, which in turn enables a need-for advice judgment (Dunning, 2011). Therefore, those with high skill should know when they have likely made an error and will seek advice, while those with low skill should be blind to their errors and thus less willing to seek advice (the Dual Burden Account; Dunning, 2011).

This analysis assesses whether metacognition is the sole source of participants' increased advice taking or whether skill and metacognitive ability work alongside one another to enhance advice taking. If metacognition is the reason why higher skilled participants are better at identifying when they have made an error and need advice, then skill will no longer be a significant predictor of advice taking. Only metacognition will significantly predict advice taking behaviors.

A multi-level model assessing accepting advice (0 = did not, 1 = did) with participants' pre-advice confidence (grand mean centered), participants' task score (grand mean centered), and participants' metacognitive sensitivity (AUC) scores (grand mean centered) and the interaction between these variables entered into the model as fixed effects. Random intercepts were included for each participant, as were random slopes based on participants' pre-advice confidence. Optimizer "bobyqa" was used. The results of this analysis are highlighted in Table 6.

Table 6.

Examining the Interaction Between Skill, AUC, and Confidence

Predictors	*Log-Odds*	*std. Error*	*p*
Intercept	-1.14	0.10	**<0.001**
Confidence [a]	-2.30	0.24	**<0.001**
Logic Score [a]	0.89	0.56	0.113
AUC [a]	1.57	0.81	0.052
Confidence X Score [a]	4.35	1.27	**<0.001**
Confidence X AUC [a]	-4.27	1.93	**0.027**
Score X AUC [a]	0.42	4.28	0.922
Score x Confidence X AUC [a]	-3.51	9.91	0.723
Random Effects			
τ_{00} (Variance of Random Intercepts)		1.31	
τ_{11} (*SD* of Random Slope)		.50	
ρ_{01} (Correlation: Intercepts & Slopes)		- 0.75	
N		224	
Number of Trials		2753	

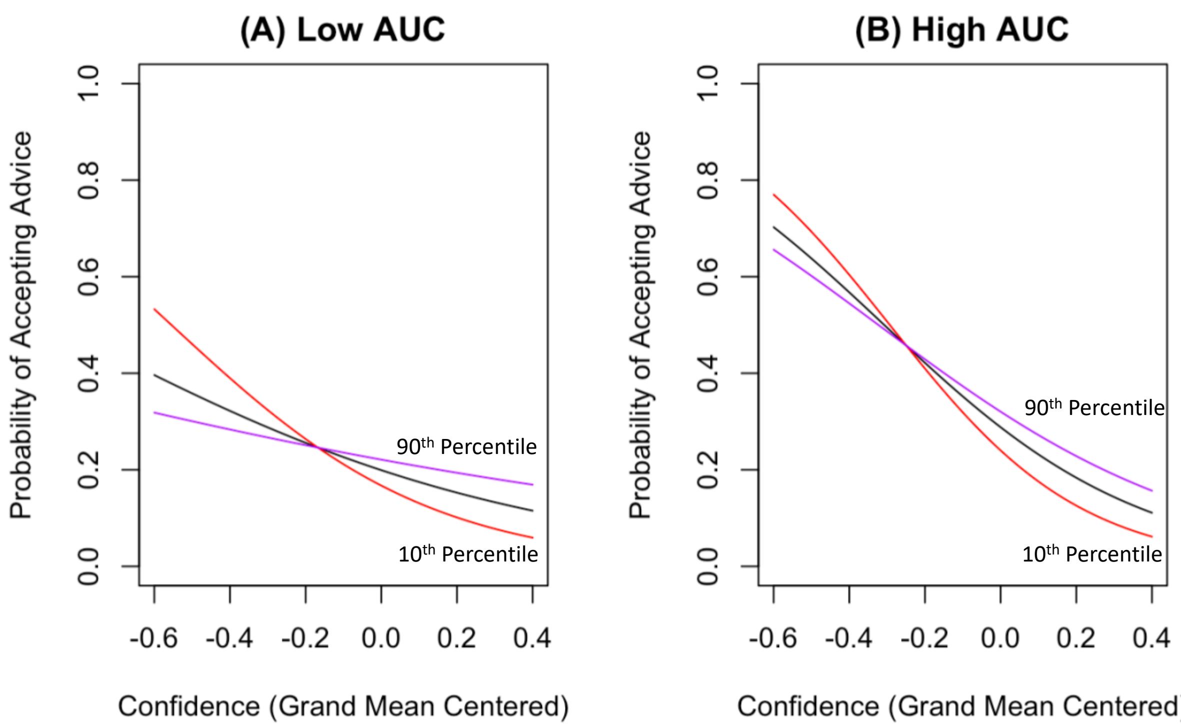
Note.
[a] Grand Mean Centered Variable(s)

Confidence was found to be a significant negative predictor of advice taking, b = -2.30, p < .001. However, neither score (b = .89, p = .113) nor AUC (b = 1.57, p = .052) were found to be significant predictors on their own. Rather, the interaction between skill and confidence and confidence and AUC were found to significantly predict advice taking patterns. This highlights that participants' confidence in their initial answer plays a key role in their decision-making regarding advice taking.

To examine advice taking patterns of low and high AUC and low and high skilled participants at different levels of confidence, the model was used to extract predicted probabilities of taking advice from individuals who scored within the 10th and 90th percentile on the logic task, as well as individuals with 10th and 90th percentile AUC values. These high and low scoring individuals' and high and low AUC individuals' advice taking patterns is explored at different pre-advice confidence levels. The results of this analysis are highlighted in Figure 20.

Figure 20.

Predicted Probability of Accepting Advice for 90th Percentile and 10th Percentile Logic Score & AUC Participants at Different Levels of Confidence



Note. The black line represents the fixed intercept and slope for the sample. The red line represents the intercept and slope of an individual who scored in the 10th percentile. The purple line represents the intercept and slope of an individual who scored in the 90th percentile.

The range of AUC values for each logic score quantile is highlighted in Table 7. To start, participants with low task scores but high AUCs were examined (Panel B, red line). These individuals scored in the 10th percentile (21% correct) but had AUCs in the 90th percentile (.82 on a scale of 0 to 1). When their confidence was low, these individuals were approximately 77% likely to accept advice. In contrast, low-scoring individuals with low AUC (10th percentile = .50 AUC), as demonstrated in Panel A, were approximately

53% likely to accept advice when their confidence was low. However, when confidence was high, low-scoring individuals with both 10^{th} and 90^{th} percentile AUC scores (the red lines on Panels A and B) were only approximately 6% likely to accept advice, respectively.

Table 7.

AUC Ranges for Each Score Quantile

Score Quantile	Mean Task Score for Quantile	AUC Range
0%	0.08	[0.50, 0.80]
10%	0.21	[0.38, 0.84]
20%	0.29	[0.41, 0.80]
30%	0.38	[0.49, 0.89]
40%	0.42	[0.51, 0.78]
50%	0.46	[0.47, 0.90]
60%	0.50	[0.51, 0.87]
70%	0.58	[0.38, 0.90]
80%	0.63	[0.41, 0.85]
90%	0.71	[0.49, 0.93]

Next, high-scoring participants were examined. High scoring individuals who had high AUCs (90^{th} percentile = .82), were approximately 66% likely to accept advice when their confidence was low (Panel B, purple line). In contrast, high-scoring, low-AUC individuals were approximately 32% likely to accept advice. On the other hand, when their confidence was high, high-scoring, low-AUC and high-scoring, high-AUC individuals were 16% to 17% likely to accept advice, respectively.

This analysis demonstrates that, for both skill groups when confidence is high, advice taking is low. However, high skilled participants are approximately 10% more likely to accept advice when their confidence is high than the low-skilled individuals, averaging across all levels of metacognition. Yet, when confidence is low, low-skilled participants of are approximately 11% to 45% more likely to accept advice than their high-skilled peers, averaged across all levels of metacognition. Therefore, confidence has more of an impact on participants' willingness to take advice, but high skill leads to slightly greater chance of using advice when in a high confidence state. This is likely due to the fact that the high-skilled have a greater ability to ascertain why the advisor's answer may be accurate when it differs from theirs.

It is important to note that while the interaction between confidence and score, as well as the interaction between confidence and AUC were significant, the three-way interaction between confidence, score, and AUC was not, b = - 3.51, p = .723. Therefore, there are not statistically significant differences between different skill-metacognitive ability pairings at different levels of confidence regarding their willingness to take advice.

These results highlight low support for the dual-burden account; the high-skilled are not the only ones with high-metacognitive ability, and thus it is not their superior metacognitive ability that allows them to make better decisions about their performance or their need for advice. Rather, both high and low-skilled individuals can make poor decisions regarding whether they should take advice when in a high-confidence state. However, the high skilled have a slightly higher likelihood of accepting advice in a high-confidence state, regardless of their metacognitive ability (~10% greater likelihood for low and high-AUC, high-skilled participants).

Which Model Best Explains the Data? Comparing the High Confidence, Skill-Based and Dual Burden Metacognition Models

In this chapter, the primary focus was to determine which of the two hypotheses, the Skill-Based or High Confidence, better explains patterns of advice-taking. Three multi-level models were constructed to investigate the effects of confidence, skill, and metacognitive sensitivity on advice-taking patterns. The first model (Model A) was designed to the test the High Confidence hypothesis, which states that when individuals are highly confident, they disregard advice, irrespective of their skill level, due to confidence-induced blindness to the need for advice. Conversely, the second model (Model B) tested the Skill-Based hypothesis, which suggests that ignoring advice when in a high-confidence state should only occur among individuals of low-skill. Additionally, the final model (Model C) considered the Dual Burden hypothesis, which posits that highly-skilled individuals are better at recognizing their errors due to greater metacognitive abilities, making them more likely to accept advice because they recognize the need for it.

To determine which model provides the best explanation for the data, these models were entered into a one-way ANOVA in R, which evaluated their structures and fit. Model AIC and Chi-Squared Tests were used to compare the models. Based on AIC alone, Models B exploring the roles of skill and confidence (AIC = 2907.6) and Model C metacognition, skill, and confidence (AIC = 2907.8), were essentially tied. Model A which only considered the role of confidence was the poorest fitting model, (AIC = 2917.8).

The Chi-squared test did correspond with this conclusion. Model B (Skill-Based) fit the data better than Model A (High Confidence), $\chi^2(2, N = 224) = 14.24$, $p < .001$. However, Model C (metacognition, skill, and confidence) did not fit better than Model B (Skill-Based), $\chi^2(4, N = 224) = 7.81$, $p = .099$. Therefore, understanding an individual's domain skill and decision-making confidence is more important than capturing their metacognitive ability. Thus, the Dual Burden assertation of better decision-making of the high-skilled due to metacognitive ability is not supported. However, the Skill-Based assertation that the high-skilled are (slightly) better at taking advice when in a high confidence state, is supported.

Chapter VI Discussion

Meeting the Assumptions of the Skill-Based and High Confidence Hypotheses

The goal of this chapter was to explore two hypotheses: the Skill-Based hypothesis and the High Confidence hypothesis. The High Confidence hypothesis states that people are especially sensitive to their internal subjective feelings of confidence in their initial answer, and relatively insensitive to external cues, such as disagreement between their and the advisor(s)' answer. Thus, they are less likely to accept advice when in a high-confidence state, as this confidence blinds them to the need for advice. The strong version of the High Confidence hypothesis assumes that error blindness is a universal feature of human cognition, irrespective of skill level. Therefore, the first part of this chapter examined whether low-skilled and high-skilled participants' confidence judgments on their error trials significantly differed.

The Skill-Based hypothesis stated that the high-skilled, due to their superior ability to identify when they are incorrect, are better able to use advice as a cue to rethink their initial answer. Therefore, the assumption that individuals with higher skill (as measured by task score) are better able to assess when they are correct versus incorrect (i.e., metacognitive sensitivity) was examined. Results, which are highlighted below, indicate that both the assumptions of the High Confidence and Skill-Based Hypotheses were empirically supported.

The High Confidence Hypothesis Assumption of High Confidence Errors. Both the expected Dunning-Kruger and Error Blindness patterns were found in this sample. Those in the bottom quartile overestimated both the percentage of the task they would answer correctly and their percentile ranking, while those in the top quartile either underestimated or accurately estimated their performance, replicating Dunning-Kruger patterns (e.g., Kruger & Dunning, 1999; Ehrlinger et al., 2008; Pennycook et al., 2017). However, both low and high-skilled individuals demonstrated roughly equally high confidence on their error trials, replicating the Universal Error Blindness Hypothesis results from Sanchez, Benson, and Ruthruff (2023). Therefore, the assumption of the High Confidence hypothesis of perpetration of high confidence errors across skill groups was supported.

The Skill-Based Hypothesis Assumption of Superior Metacognitive Ability in the High-Skilled. Examining participants' metacognitive sensitivity, a skill effect was found. Participants in Quartile 4 (top 25% of participants) had higher AUC values than their Quartile 1 peers (bottom 25%) of participants. In addition, regression analyses highlighted a positive relationship between task score and skill. Thus, the Skill-Based hypothesis assumption of greater metacognitive sensitivity in higher skilled individuals was empirically supported.

The High-Skilled Are More Likely to Take Advice When in a High Confidence State, But Only by a Narrow Margin

To test the High Confidence and Skill-Based hypotheses, multi-level logistic regressions were used to consider individual variability. Results indicated that those with higher task skill (i.e., 90th percentile) were more likely to take advice when their pre-advice confidence was high than those who had lower skill (i.e., 10th percentile), supporting the Skill-Based hypothesis. Therefore, the high degree of blindness to advice in all skill groups predicted by the High Confidence hypothesis was not seen.

It is important to note that while the probability of taking advice in a high-confidence state is higher in high-skilled individuals than their lower-skilled peers, this probability is still extremely low. The high-skilled had an approximately 18% probability of taking advice, while the low-skilled had an approximately 7% likelihood of taking advice while in a high confidence state. Overall, the low probability of taking advice by either skill group will likely have consequences in high-stakes domains for which a single error can have grave consequences.

In addition, those with high metacognitive sensitivity were not able to catch and correct all of their errors. The Skill-Based hypothesis predicted that those with high skill would have a better ability to detect when they were likely to incorrect, which is measured by participants' metacognitive sensitivity (AUC). High AUC individuals were great at taking advice when their pre-advice confidence was low, with the probability of taking advice ranging from approximately 77% in the low-skilled and 66% in the high-skilled. In low-AUC individuals, the low-skilled took advice 53% of the time, while the high-skilled took advice 32% of the time.

However, when pre-advice confidence was high, the probability of taking advice decreases sharply. 10^{th} percentile scoring participants were only approximately 6% likely to accept advice across all metacognition levels, while 90^{th} percentile participants were only approximately 16 – 17% likely to accept advice, averaging across all levels of metacognitive ability. Therefore, one interpretation of this result is that some errors are "slipping past" participants' conscious awareness, even when they have superior metacognitive discrimination. They do not get a signal that they have made an error when in a high confidence state and thus are unable to update their prior beliefs when presented with advice. This supports the Universal Error Blindness Hypothesis assertion that error blindness is not due to deficits of domain skill or poor metacognitive ability, but rather due to blindness induced by high confidence.

Consequently, rather than treating the Skill-Based and High Confidence hypotheses as competing hypothesis, the results of this investigation highlight that both hypotheses have merit and should be combined. People are largely error blind, and this makes them less likely to take advice, even amongst the high-skilled. However, there is a small effect of skill on advice taking, with the high-skilled taking slightly more advice than the low-skilled when in a high-confidence state. This increase in advice taking may be due to a high-skilled individual's greater ability to successfully identify why the advised answer is correct. Meanwhile, a low-skilled individual is unable to reason why the advised answer is correct and thus rejects it. Therefore, increasing low-skilled participants' skill in a domain, may increase advice taking when in a high-confidence state.

However, it is interesting to note that, when in a low-confidence state, the low-skilled take advice more often than their high-skilled peers. This may be due to the fact that the low-skilled are conditioned to need more assistance from others, so when they can detect that they are incorrect, they are more receptive to assistance from others.

Chapter VI: Conclusion

Results of this chapter highlight that both the High Confidence and Skill-Based hypotheses are important in explaining participants' advice taking behavior. In terms of the Skill-Based hypothesis, skill predicted participants' ability to estimate their performance, their metacognitive sensitivity, and their advice taking behavior. However, while skill was found to predict greater advice taking, this effect was significantly moderated by participants' pre-advice state of confidence.

When high-skilled participants are in a high-confidence state, they are likely to take advice on only one out of every six errors they make (~ 18%), while the low skilled are likely to take advice on one out of every fourteen errors they make (~ 7 %). Therefore, when individuals of all skill levels are highly confident in their initial answer, they are likely to ignore advice 82% to 93% of their errors. Thus, those who are highly confident in their performance are often blind to their errors, consistent with a moderate version of the High Confidence hypothesis. However, individuals with higher skill are better able to use advice to determine why the advised answer is correct, which leads to the slightly higher advice taking rates in this population.

Chapter VII: Individual Difference Investigation Results

The ultimate question this chapter investigations is *who* is most (or least) likely to accept advice, a question largely neglected in the advice-taking literature. Yet, answering this question is critical for many domains for which catching one's errors is essential, such as emergency response, finance, aviation, and cyber security. Identification of individuals who are more likely to accept advice when they have made an error can enhance personnel selection in these high-stakes domains.

Prior research has indicated that there are large individual differences in advice taking patterns, with some individuals never taking advice and others taking advice nearly three-fourths of the time (e.g., Pescetelli, Hauperich, and Yeung, 2021). However, the literature has yet to fully explain why these differences in advice taking exist. Thus, the current investigation aims to fill this gap through an in-depth investigation of the roles of personality, metacognitive ability, and domain skill play in enhancing or inhibiting advice taking.

Individuals who are high in the personality traits of Conscientiousness, Agreeableness, Extraversion, Neuroticism, and normative social influence are predicted to be more receptive to advice, while those who were high in Openness are predicted to be less receptive. An additional feature of this empirical investigation includes examining confidence, metacognitive sensitivity, skill, and advisor preference as individual differences.

Skill is not a typically considered an essential factor in advice taking investigations. However, as highlighted in *Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results*, skill can shape participants' ability to accurately assess their performance and their advice taking patterns. Therefore, the role of skill in advice taking behaviors is considered an essential element to not only measure, but also as an avenue to potentially enhance advice taking.

Participant Characteristics

50.9% (N = 114) of the sample collected identified themselves as female, 47.3% (N = 106) of the sample identified as male, and 1.8% (N = 4) of the sample identified as non-binary. The mean age of the sample was M = 38.08 years of age (SD = 13.51). The minimum age was 18 years of age, and the maximum was 78 years of age.

Correlations By Advisor Type

Correlations between participants' advice taking behavior and individual differences of interest can be found in Table 8. As many participants did not take advice, percentage of advice use on one's error trials is considered zero-inflated data. Zero-inflated data violates the assumptions of parametric correlation techniques and Kendall's tau or Spearman's Rho should be used (Pimentel, Niewiadomska-Bugaj, & Wang, 2015). R package Mazeinda was used to calculate the correlations between percentage of advice used[4] and the personality, demographic, metacognition, and skill variables (Dolma Albasi & Wells, 2022).

[4] Percentage of advice use was calculated using the following equation: *Number of Trials Advice was Taken* divided by *the Total Number of Errors* made on the logic task for each participant.

Table 8.

Correlations Between Advice Taken (%) and Personality, Performance, & Demographic Predictors

Variable	Kendall's τ with Advice Use %
Task Score (%)	0.11*
Age	-0.06
Extraversion	0.02
Openness	0.03
Conscientiousness	-0.01
Agreeableness	0.04
Neuroticism	0.03
Normative Social Influence	0.12*
AUC	0.14**

Note. * Indicates $p < .05$. ** indicates $p < .01$. These correlations are between uncentered variables.

Results indicate that those with higher scores on the logic task were more likely to use advice, $r_\tau = .11$, $p = .02$. Individuals' susceptibility to normative social influence correlated with greater use of advice, $r_\tau = .12$, $p = .01$. AUC was also positively correlated with advice use, $r_\tau = .14$, $p < .001$. No other significant correlations between percentage of advice use and the personality, demographic, or personality variables were found.

As the personality, demographic, metacognition, and skill variables are not zero-inflated and follow a normal distribution, Pearson product correlations were used to assess relationships between these independent variables. These can be found in Table 9. Table 9 does not contain correlations between percentage of advice use (dependent variable) and the demographic, metacognition, and skill variables (independent variables) due to the zero-inflation problem – see Table 8 for the correlations between the dependent and independent variables.

Table 9.

Means, Standard Deviations, and Correlations with Confidence Intervals of Predictor Variables

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Task Score (%)	0.47	0.18								
2. Age	38.08	13.51	-0.01							
			[-.14, .12]							
3. Extraversion	22.65	7.85	-0.05	0.11						
			[-.18, .08]	[-.02, .24]						
4. Openness	36.46	7.37	-0.02	0.1	**.30****					
			[-.15, .11]	[-.03, .23]	[.17, .41]					
5. Conscientiousness	34.36	7.57	-0.02	**.23****	**.41****	**.20****				
			[-.15, .12]	[.10, .35]	[.29, .51]	[.07, .33]				
6. Neuroticism	22	8.72	-0.06	**-.21****	**-.48****	**-.22****	**-.65****			
			[-.19, .07]	[-.33, -.08]	[-.57, -.37]	[-.34, -.09]	[-.72, -.57]			
7. Agreeableness	34.21	6.99	-0.05	**.23****	**.30****	**.31****	**.49****	**-.58****		
			[-.18, .08]	[.10, .35]	[.17, .41]	[.18, .42]	[.38, .58]	[-.66, -.48]		
8. Normative	7.88	4.01	0.02	**-.17***	0.11	-0.06	0.01	-0.06	0.02	
			[-.11, .15]	[-.29, -.04]	[-.03, .23]	[-.19, .07]	[-.12, .14]	[-.19, .07]	[-.11, .15]	
9. AUC	0.67	0.12	**.27****	0.04	-0.1	0.06	0.04	-0.06	-0.02	-0.03
			[.14, .38]	[-.09, .17]	[-.23, .03]	[-.08, .19]	[-.09, .17]	[-.19, .07]	[-.15, .12]	[-.16, .10]

Note. These correlations are between uncentered variables.

The Role of Personality in Advice Taking

A multi-level logistic regression model was constructed to examine the effects of age, gender, the Big-5 traits, normative social influence, task score, and AUC on the probability of accepting advice. Prior research has indicated that participants' age and gender do not differentially predict advice taking (Bailey et al., 2022). Thus, participants' age and gender were entered into the model as control factors but are not considered predictors of interest.

$$\boldsymbol{p(Switching\ to\ Advisor's\ Answer)_{ij}} =$$

$$\beta_{0j} + \beta_1 Age_i + \beta_2 Gender\ Identity_j + \beta_3 Conscientiousness_j + \beta_4 Agreeableness_j + \beta_5 Neuroticism_j +$$

$$\beta_6 Openness_j + \beta_7 Extraversion_j + \beta_8 Normative\ Influence_j + \beta_9 AUC_j +$$

$$\beta_{10} Task\ Score_j + \beta_{11j} Pre - advice\ Confidence_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{11j} = \beta_{11} + u_{1j}$$

where *i* represents each logic question nested within participant *j.* β_{0j} represents the random intercept for each participant, while β_{11j} represents the random slope of pre-advice confidence for each participant *j*. All models were implemented using the glmer function of the lme4 package in R (Bates, 2021) and were fitted with restricted maximum likelihood. Please note that only error trials were used for this analysis (i.e., the correct trials were removed from analysis). When advice was 100% correct, advice could not realistically be considered "taken" on correct trials.

Model Construction

The construction of the multi-level model was informed by the process outlined by Sommet and Morselli (2017). First, the personality measures, task score, AUC, confidence, and age were grand mean centered (i.e., centered based on the mean for the whole sample). Gender identity was simple effects coded. Participant ID served as the clustering variable.

Next, a baseline model was constructed to examine the degree of variance across participants regarding the odds that they would accept the offered advice. This baseline model included no predictors and had random intercepts for each participant.

After the baseline model, a random intercept model was constructed with the demographic, personality, confidence, metacognition, and performance predictors entered into the model. Finally, a random intercept and slope model was constructed with pre-advice confidence as the random slope across participants. The optimizer bobyqa was used for the random intercept and random intercept and slope models. Both the random intercept and random intercept and slope models were assessed for multicollinearity and no violations were found (VIF ranges: 1.00 – 2.52 and Tolerance ranges: .40 to 1.00).

Model Comparison & Selection

The random intercept and random intercept and slope models were compared to one another. AIC values indicate that the random intercept model (AIC of 2913.0) explained the data better than the random intercept and slope model (AIC of 2916.1). However, a Chi-square test failed to detect a discernable difference between models, $\chi^2(2, N = 224) = .88$, $p = .644$. Therefore, the random slope model was selected for interpretation, as the role of confidence increasing or decreasing the likelihood of accepting advice is of interest as an individual difference in this investigation.

Model Overview

The results of the random intercept and slope models are highlighted in Table 10. Due to the scaling issue with odds ratio, log-odds were reported.

Table 10.

Effects of Hypothesized Predictors on Likelihood of Taking Advice

	Random Intercept Model			**Random Intercept & Slope Model**		
Predictors	*Log-odds*	*std. Error*	*p*	*Log-odds*	*std. Error*	*p*
Intercept	-1.40	0.14	**< 0.001**	-1.38	0.15	**< 0.001**
Age[a]	-0.01	0.01	0.374	-0.01	0.01	0.379
Gender: Male	0.17	0.20	0.406	0.16	0.21	0.443
Gender: Non-Binary	0.06	0.72	0.938	0.05	0.74	0.945
Conscientiousness[a]	0.01	0.02	0.507	0.01	0.02	0.604
Agreeableness[a]	0.03	0.02	0.111	0.03	0.02	0.113
Neuroticism[a]	0.02	0.02	0.334	0.02	0.02	0.345
Openness[a]	0.01	0.01	0.427	0.01	0.01	0.386
Extraversion[a]	0.02	0.01	0.273	0.01	0.01	0.316
Normative Influence[a]	0.06	0.02	**0.014**	0.06	0.02	**0.017**
AUC[a]	1.71	0.81	**0.034**	1.61	0.83	0.053†
Logic Score[a]	0.90	0.54	0.096†	1.00	0.56	0.077†
Pre-Advice Confidence[a]	-2.61	0.21	**< 0.001**	-2.61	0.23	**< 0.001**
Random Effects						
τ_{00} (Variance of Random Intercepts)	1.24			1.20		
τ_{11} (*SD* of Random Slope)				0.79		
ρ_{01} (Correlation: Intercepts & Slopes)				-0.16		
N	224			224		
Number of Trials	2753			2753		

Note.
[a] Grand Mean Centered Variable
† Indicates $p < .10$

Random Effects. The fixed intercept indicated that participants were unlikely to take advice when confidence, demographics, personality, and performance variables were at the sample mean, b = -1.38, se = .15, p < .001. However, participants differed in their willingness to accept advice, with random intercepts indicating a standard deviation of 1.10 in log-odds to the fixed intercept of -1.38 log-odds. Therefore, consistent with earlier analyses, significant individual differences were found in participants' willingness to accept advice.

Pre-advice confidence on one's error trials also played a substantial role in inhibiting participants' willingness to accept advice. From the fixed slope of - 2.61 log-odds, participants had a slope standard deviation of .79. The random slope and random intercept were also negatively correlated, r = - .16, which indicates that higher individual confidence on error trials led to lower probability of taking advice. Taken together, the evidence highlights that some individuals are highly resistant to advice when they are in a high-confidence, but erroneous state.

Fixed Effects: Age & Gender. Consistent with previous literature (Bailey et al., 2022), gender was not found to be significant predictors of accepting advice, as highlighted in Table 10. Males were not less likely to take advice than females, b = .16, se = .21, p = .443. While consistent with the advice taking literature, this result is surprising from an overconfidence perspective. Males tend to be more overconfident than females (Moore & Dev, 2017; See et al., 2011), and high-confidence states have been found to lead to low advice taking in previous chapters (i.e., approximately 7% - 17%). In addition, age was not found to increase advice use, b = - .01, se = .01, p = .379. See Table 11 for the percentage of the sample in each age range.

Table 11.

Distribution of Ages in the Sample

Age Range	Number of participants	Percentage
Under 20	4	1.8%
20s	62	27.7%
30s	78	34.8%
40s	33	14.7%
50s	26	11.6%
60s	15	6.7%
70s	6	2.7%

Fixed Effects: Big-5. Hypotheses stated that those high in the traits of Agreeableness, Conscientiousness, Neuroticism, and Extraversion would be more willing to accept advice, while those high in Openness would be less willing. These predictions were not supported, with these traits <u>not</u> found to predict the probability of accepting advice at the $p < .05$ level, as highlighted in Table 10.

***Fixed Effects: Susceptibility to Persuasion*.** The hypothesized relationship between normative social influence and advice taking is that individuals high in normative social influence would accept more advice due to their desire to be perceived as acting in accordance with how others behaved. As hypothesized, normative influence was found to be a significant predictor of advice acceptance, $b = 0.06$, $se = 0.02$, $p = .017$.

Fixed effects: Confidence & Performance. Task score had a surprising relationship with the probability of accepting advice. Per the Skill-Based hypothesis, explored in the last chapter, skill was expected to lead to a greater probability of accepting advice. Task score had a positive relationship with the probability of accepting advice, $b = 1.00$, $se = 0.56$, $p = .077$, but this relationship was not statistically significant.

In *Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results*, a significant interaction between confidence and skill was identified. Therefore, participants' pre-advice confidence may render the main effect of score non-significant, as pre-advice confidence is a strong moderator of the relationship between logic score and advice taking.

Similarly, metacognitive sensitivity (AUC) was expected to have a positive relationship with advice acceptance. A positive relationship was found between AUC and advice taking but failed to reach statistical significance in the random intercept and slope model, b = 1.61, se = .83, p = .053. It is important to note that the relationship between AUC and advice taking did reach significance in the intercept only model, b = 1.71, se = .81, p = .034. Therefore, the relationship between confidence and advice taking as captured by the random slopes may be a better predictor of advice taking patterns than AUC calculated through Receive Operator Curves.

In addition, consistent with predictions, pre-advice confidence had a strong negative relationship with probability of accepting advice, b = - 2.61, se = .23, p < .001. Therefore, those individuals who were overconfident on their error trials were the least likely to accept advice.

Chapter VII Discussion

This investigation was successful in categorizing who is most likely to accept advice, with considerable variation seen across individuals in their probability of accepting advice. Random intercepts highlighted that some individuals are likely to never accept advice, while others are likely to regularly accept advice. Furthermore, this willingness to accept advice is modulated by susceptibility to normative social influence, and the perpetration of high confidence errors.

Consistent with predictions, those who are high in the personality trait of normative social influence were more likely to accept advice. Those high in normative social influence are concerned with appropriate ways to act in a social situation. Thus, they will acquiesce with others' decisions to be in accordance with the perceived social rules governing a situation. Therefore, when considering personnel for roles that involve accepting recommendations from others or critical systems, those high in normative social influence would make excellent candidates.

Confidence also served as a strong negative predictor of advice taking, with high confidence on one's error trials leading to low probability of accepting advice. Therefore, those who are overconfident in their abilities and are unable to detect when they have made an error (i.e., perpetrating a high confidence error) are more likely to not take advice when they need it. It is important to note that committing high confidence errors was seen across all skill levels and metacognitive abilities.

As explored in Chapter VI: Metacognition, the Dunning-Kruger Effect, & Confidence Results, having high metacognitive sensitivity did not protect against high confidence errors, with individuals in highly confident states having an approximately 6% (Low-Skill, Low & High AUC) to 18% (High-Skill, Low & High AUC) probability of accepting advice. Thus, while the results in this investigation point to ways to enhance advice use by selecting individuals who are more susceptible to normative social influence, high confidence errors pose a serious threat to advice taking in domains where a high degree of advice taking is necessary (e.g., emergency and disaster response and responding to cyber threats).

Chapter VII: Conclusion

Significant individual variation in the probability of taking advice on one's error trials was found in this sample, with some individuals never accepting advice while others took a substantial amount of advice. The personality trait of susceptibility to normative social influence were found to lead to greater advice taking. However, participants' confidence was found to be a significant inhibiter of advice taking, with high pre-advice confidence leading to lower advice taking. However, significant individual differences were also found in the relationship between confidence and the probability of accepting advice, with some individuals being better able to accept advice when in a high confidence state.

Chapter VIII: General Discussion

Many of us can recount instances of family, friends, or even strangers who fail to heed advice when they should. These individuals proceed along a risky path only to be surprised when they experience negative consequences from their actions. The enduring question is why these individuals fail to heed well meaning, accurate advice? Thus, the primary goal of the present work was to ascertain why some individuals rarely accept advice (even when they should) and why others are better at using advice to correct their mistakes.

This work addressed three novel questions related to individual differences in advice taking. First, could the source of the advice (crowd versus a single advisor) be leveraged to enhance advice utilization when the advice is 100% accurate (the Social Consensus hypothesis)? Second, does participants' skill level and metacognitive ability predict greater advice utilization (Skill-Based Advice Taking hypothesis) or does high confidence on one's incorrect trials blind individuals of all skill levels to the need for advice, leading to poor advice utilization (the High Confidence Leads to Feedback Blindness hypothesis)? Finally, can personality (Conscientiousness, Agreeableness, Neuroticism, Extraversion, and Openness and normative social influence), performance (task skill and metacognitive ability), metacognitive sensitivity, and pre-advice confidence (degree of high confidence errors) factors be used to profile who is most and least likely to accept advice? The results of the empirical investigation into these questions and their unique contribution to the literature is highlighted below.

Testing the Social Consensus Hypothesis: The Crowd Does Not Influence Advice Taking to a Greater Extent Than a Single Advisor

Previous research has consistently found advice utilization rates hovering between 20% and approximately 30% (e.g., Harry & Fischer, 1997; Soll & Larrick, 1999; Soll & Larrick, 2009; Duan, Gu, & Sun, 2016; Pescetelli, Hauperich, & Yeung, 2021). This occurs even when the advisors are highly accurate (75% to 80%) or in high consequence domains (e.g., Harry & Fischer, 1997; Pescetelli, Hauperich, & Yeung, 2021). However, even when using 100% advice, as was done in this study, participants had a 25.6%, 95% CI [.22 – .30] probability of accepting advice on average. Clearly, the accuracy of the advice is insufficient to lead to greater advice utilization alone.

Thus, one central question of this investigation was the role multiple advisors (i.e., a "crowd") would play in enhancing advice use. Advice taking paradigms typically only use a single advisor to provide advice to participants. However, social conformity literature has highlighted the role that multiple recommenders can play in increasing advice utilization. Prior to this investigation, the timing of advice presentation prevented direct comparison between advice taking and social conformity research. Therefore, this study was the first to explore the relationship between greater advice use from a single versus a crowd of advisors when advice was 100% accurate.

Despite 100% accuracy of the advice, participants resisted taking advice on their error trials, utilizing advice only had a 25.6% probability of accepting advice. This effect is in-line with previous research findings of low advice utilization. While the crowd manipulation was hypothesized to be a solution to enhancing advice, per the Social Consensus hypothesis, the crowd advisor was not used significantly more than the single advisor, with the probability of using from either advisor type remaining extremely low (24.8% for the single advisor, 95% CI [.209 – .290] and 25.7% for the crowd advisor, 95% CI [.22 – .30]).

In addition, while participants with high scores on the task were approximately 30% more accurate in assessing the advisors' capabilities than those with low task scores, they were still reticent to use advice from either advisor. The lack of advice taking on error trials from both the high-skilled and low-skilled was hypothesized to be due to high confidence on these participant's error trials (i.e., the High Confidence Leads to Feedback Blindness hypothesis). Therefore, while high-skilled participants perceive that the single and crowd advisor are more accurate than the low-skilled individuals do, they often fail to receive a signal that they are incorrect. Therefore, they persist with their initial answer on the logic problem, attributing the conflict between their and the advisor's answer to the advisor being mistaken. Therefore, confidence on one's error trials is an important factor in the relationship between participants skill (as measured by task score), perceptions of advisor accuracy, and advice use.

The implication of these results is that while participants' trust the crowd slightly more and believe them to be slightly more accurate, presenting advice from a crowd or designing recommendation systems that leverage social influence elements is unlikely to substantially increase advice taking. Therefore, the onus for low advice use is less about the advisor (single or crowd) and the advice itself (which was 100% accurate), but rather has more to do with cognitive and individual factors of the judge that lead them to misestimate the abilities of the advisor and themselves. In addition, individual differences were found in participants' preference for one advisor or the other, with some preferring the single advisor over the crowd advisor and vice versa. Therefore, designers of recommendation systems should keep individual preferences in mind when building systems designed for specific individuals.

Skill Matters in Advice Taking, But High Confidence Matters More

The second novel feature of this study explored the interaction between participants' skill (as measured by their task score), their ability to determine when they are correct and incorrect (metacognitive sensitivity), and decision-making confidence in advice use. Prior advice taking research had largely ignored the role of skill and metacognitive ability in advice taking, despite other literatures finding large overconfidence effects in individuals' ability to accurately assess their capabilities (e.g., Kruger & Dunning, 1999). Overestimating one's ability was likely to hinder individuals' advice use, as they will not be able to accurately identify when they are incorrect and in need of help (i.e., poor metacognitive sensitivity). Skill was expected to directly impact participants' metacognitive sensitivity, as skill allows individuals to accurately assess differences between their correct and incorrect trials (i.e., the Dual Burden hypothesis; Dunning, 2011).

Two hypotheses were considered in this work: the High Confidence Leads to Advice Blindness hypothesis and the Skill-Based Advice Taking hypothesis. The High Confidence Leads to Advice Blindness hypothesis is based on prior research by Sanchez, Benson, and Ruthruff (2023), which has highlighted that regardless of participants domain skill, when in a high confidence state, all individuals are blind to their errors (i.e., Universal Error Blindness hypothesis). The High Confidence Leads to Advice Blindness hypothesis states that people are especially sensitive to their internal subjective feelings of confidence in their initial answer, and relatively insensitive to external cues, such as disagreement between their and the advisor(s)' answer. Thus, they are unlikely to accept advice when in a high-confidence state, as this confidence blinds them to the need for advice. The strong version of this High Confidence hypothesis assumes that this advice blindness is a general feature of human cognition, applying to people of all skill levels.

The Skill-Based Advice Taking hypothesis is based on the Dual Burden account by Dunning (2011). The Skill-Based Advice Taking hypothesis states that, due to their superior ability to identify when they are incorrect (i.e., metacognitive sensitivity), the high-skilled are better able to use advice as a cue to rethink their initial answer. Thus, even when they are initially in a high-confidence state, they are more likely to use advice than are those with low skill.

Results indicated that those with higher task skill were more likely to take advice when their pre-advice confidence was high than those with lower skill, providing empirical support for the Skill-Based hypothesis. However, while the probability of taking advice in a high-confidence state is higher in high-skilled individuals than their lower-skilled peers, these probabilities are still extremely low. The high-skilled had an approximately 18% probability of taking advice, while the low-skilled had a 7% likelihood of taking advice on their high confidence error trials.

This work also tested the assertion that the high-skilled are more likely to take advice due to their superior metacognitive ability. Results indicated that while metacognitive ability is an important factor in advice taking, with higher metacognition leading to greater advice taking, this result was not contingent on domain skill. Rather both low and high-scoring participants with high metacognitive ability were more willing to take advice. However, confidence was a stronger determining factor in advice taking patterns. When in a high confidence state, participants were less willing to take advice, regardless of their skill level or metacognitive ability.

Some Errors are Likely to Slip Past Individuals, No Matter How Skilled They Are

Therefore, one interpretation of this result is that some errors are "slipping past" participants' conscious awareness, even when they have superior metacognitive discrimination. They do not get a signal that they have made an error when in a high confidence state and thus are unable to update their prior beliefs when presented with advice. This supports the Universal Error Blindness Hypothesis assertion that error blindness is not due to deficits of domain skill or poor metacognitive ability, but rather due blindness induced by high confidence and is seen in people of all skill levels.

Consequently, rather than treating the Skill-Based and High Confidence hypotheses as competing hypothesis, the results of this investigation highlight that both hypotheses have merit and should be combined. People are largely error blind, and this makes them less likely to take advice, even amongst the high-skilled. However, there is a small effect of skill on advice taking, with the high-skilled taking slightly more advice than the low-skilled when in a high-confidence state. This increase in advice taking may be due to a high-skilled individual's greater ability to successfully identify why the advised answer is correct. Meanwhile, a low-skilled individual is unable to reason why the advised answer is correct and thus rejects it.

Implications of these findings are that while increasing domain skill can help participants make less errors and potentially catch some of their errors through the use of advice, interventions may not lead participants to catch and correct *every* error. This indicates that in domains where *any* error can have disastrous consequences, individuals in a high confidence state will fail to take advice and will proceed with actions they should not.

Who Takes Advice? The Social Normative Ones

The final novel contribution of this research is the examination of individual differences in advice taking. This work examined advice taking based on *preference for a specific type of advisor* (e.g., single versus multiple advisors), *personality* factors (Conscientiousness, Agreeableness, Neuroticism, Extraversion, and Openness and normative social influence), *skill* (task score), *metacognitive sensitivity*, and *decision-making confidence*.

This investigation was successful in categorizing who is most likely to accept advice, with considerable variation seen across individuals in their probability of accepting advice. Random intercepts highlighted that some individuals are likely to almost never accept advice, while others are likely to regularly accept advice. Random slopes indicated that participants' initial confidence in their accuracy moderates the probability of accepting advice quite significantly across individuals. Furthermore, willingness to accept advice is modulated by personality and performance factors, including susceptibility to normative social influence and the perpetration of high confidence errors.

Consistent with predictions, those who are high in the personality trait of normative social influence were more likely to accept advice. This may be due to the fact that these individuals are more likely to follow the decision-making of others out of the desire to be accepted. Thus, it is less important that those high in normative social influence are able to determine why their answer was incorrect and why the advisor's answer is accurate, but rather that they and the advisor agree.

Confidence also served as a strong negative predictor of advice taking, with high confidence on one's error trials leading to low probability of accepting advice. Therefore, those who are overconfident in their abilities and are unable to detect when they have made an error (i.e., perpetrating a high confidence error) are more likely to not take advice when they need it.

The implications of these results are that when selecting for roles that would benefit from high degree of advice taking, selecting those who score high normative social influence will lead to greater advice use. However, it is important to keep in mind that high confidence errors are still a strong preventative factor of advice taking. Therefore, perfect advice use is unlikely, even in those with high in normative social influence.

Limitations

The first limitation of this work is the inability to calculate bias-free estimates of metacognitive ability. As discussed earlier in this paper, metacognitive sensitivity calculations such as ROC curves and AUC are biased by participants' d-prime scores, which makes participants with higher task scores appear to have better metacognition simply by the virtue of scoring better on the task. However, better methods for calculating metacognitive sensitivity does not yet exist (Vuorre & Metcalf, 2021). Therefore, this work could not provide a perfect test of the relationship between task skill and AUC, with some inflation likely to exist in the high-skilled's calculated AUC. This is a limitation of the current work and of the field at large.

Second, the advice presented to participants lacked two elements that could have additionally shaped their decision-making regarding the advisor. First, participants were not provided the accuracy of the advisors in the form of feedback regarding what the correct answer was after each logic question. Prior research has shown that advisor expertise can increase advice taking by as much as 16% (32% use of low accuracy advisors to 48% use of high accuracy advisors; Bailey et al., 2021). However, providing feedback regarding the advisors' accuracy would likely negatively impact participants' post-task performance judgments, as well as their pre-advice confidence judgments. As these performance judgements were critical features of the current experiment, feedback regarding the correct answer for each logic question was not included as a feature of this work. Future work should examine the role of individual differences on accepting advice from advisors of different skill levels.

Third, the size of the crowd advisor was not specified. While participants were told that they were being displayed answers from a group of prior participants, they were not told that the advice was based on answers from a group of 268 prior participants from Sanchez, Benson, and Ruthruff (2023). Mannes (2009) has demonstrated that group size may lead to nuances in advice taking, with larger groups leading to *less* advice utilization. While low advice accuracy may have confounded Mannes's results, future work should examine the role different group sizes play in advice utilization.

In addition, a critical examination in this work is the comparison between the low-skilled participants and the high-skilled participants. While this comparison has real-world implications, with skill representing a mechanism for enhancing advice use, it is important to note that there is a limitation in the ability to compare between these groups. Individuals who are low-skilled made 17.75 (SD = 1.69) errors on average (Quartile 1) while the high-skilled made only 6.73 (SD = 1.71) errors on average. Therefore, the low-skilled participants had a greater opportunity to take advice than the high-skilled by nature of their performance, though this was not found to significantly impact the probability that individuals would accept advice (see Figure 6).

Future Work

Three theories have been posed to explain advice taking. The first is the anchoring and adjustment theory, which states that individuals settle on a decision and have trouble incorporating new information due to cognitive dissonance. Thus, to eliminate cognitive dissonance provoked by the advisor's recommendation conflicting with theirs, individuals disregard the advice. *Consistency theory* argues that individuals make decisions that are consistent with their internal attitudes, values, and personal histories (Russo et al., 1996; Simon & Holyoak, 2002). When presented information that conflicts with an individual's prior decisions, the resulting *cognitive dissonance* is resolved through *cognitive distortion* for which individuals attend to information that matches their initial belief (Russo et al., 1996; Simon & Holyoak, 2002). Therefore, the desire to maintain cognitive consistency acts as an *anchoring and adjustment* function for which individuals settle on a decision and have trouble incorporating new information. Thus, to eliminate cognitive dissonance provoked by the advisor's recommendation conflicting with theirs, individuals disregard the advice, regardless of how useful it is.

The second prevailing theory is the *self/other effect*, in which a decision-maker has access to the evidence they used to make their decision but not the evidence the advisor used (Yaniv, 2004). Due to this lack of clarity regarding others' judgments, individuals reject the advice in favor of their own judgment. The final theory is the *egocentric bias* theory, which stipulates that individuals prefer their judgment over others' judgments due to the belief that their decision-making capabilities are superior (Krueger, 2003; Bonaccio & Dalal, 2006).

The current work is most applicable to testing the egocentric bias account, as participants' self-belief and belief in the advisor was directly captured. However, examination of the *self/other effect* is scant in the literature. An extension of the current work is to examine the self/other effect by examining advice taking when the advisor's reasoning is provided to participants. To test this phenomenon, advice would be presented from two advisors: a poor advisor who provides incorrect logic and a good advisor who provides correct logic.

To test this, prior work conducted by Sanchez and Ruthruff would be leveraged. This prior work used a think aloud protocol to examine the Dunning-Kruger effect on the logic problems used in this current work. Therefore, the reasoning participants provided for each logic question can be used to create arguments for a fictitious poor quality and a fictitious good quality advisor. In this paradigm, participants will be asked to answer the logic questions and provide a pre-advice confidence judgment. Next, participants will be provided advice from the poor and good quality advisors, alongside the reasoning for how the advisor arrived at their answer. Participants will not be told which advisor is of poor and good quality, only that the quality of the advice from the two advisors differ. Participants will then be asked whether they would like to change their answer and provide an updated confidence judgment. A control condition will provide advice from two advisors without the reasoning as to why the answers are correct.

This new experiment will allow for a direct test of the self/other effect versus egocentric bias accounts, as participants will be asked to estimate their abilities and belief in their accuracy, in conjunction to being provided advisors reasons for providing specific advice.

References

Albasi Dolma, A., & Wells, M., (2022). Mazeinda: Monotonic Association on Zero-Inflated Data. https://cran.r-project.org/web/packages/mazeinda/vignettes/my-vignette.pdf

Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In L. B. Alloy (Ed.), *Cognitive processes in depression* (pp. 223–265). The Guilford Press.

Al Rashdan, A. Y., & Mortenson, T. J. (2018). *Automation Technologies Impact on the Work Process of Nuclear Power Plants* (No. INL/EXT-18-51457-Rev000). Idaho National Lab.(INL), Idaho Falls, ID (United States).

Amazeen, M. A., & Bucy, E. P. (2019). Conferring resistance to digital disinformation: The inoculating influence of procedural news knowledge. *Journal of Broadcasting & Electronic Media*, *63*(3), 415-432.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1–70. https://doi.org/10.1037/h0093718

Ashton, M.C. and Lee, K. (2001), "A theoretical basis for the major dimensions of personality", *European Journal of Personality*, Vol. 15 No. 5, pp. 327-53.

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general*, *149*(8), 1608.

Bailey, P. E., Ebner, N. C., Moustafa, A. A., Phillips, J. R., Leon, T., & Weidemann, G. (2021). The weight of advice in older age. *Decision*, *8*(2), 123–132. https://doi.org/10.1037/dec0000138

Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology*, 1-26. https://doi.org/10.1007/s12144-022-03573-2

Bajo, M. T., Gómez-Ariza, C. J., & Marful, A. (2021). Inhibitory Control of Information in Memory across Domains. *Current Directions in Psychological Science*, *30*(5), 444–453. https://doi.org/10.1177/09637214211039857

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and assessment*, *9*(1-2), 9-30.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Bates, D. (2021, June 22). *LME4: Linear mixed-effects models using 'Eigen' and S4. Linear Mixed-Effects Models using 'Eigen' and S4*. Retrieved December 4, 2021, from https://cran.r-project.org/web/packages/lme4/lme4.pdf.

Beri, D., & Redy, D. K. J. (2022). A Cognitive Insight into Political Minds: Exploring Cognitive Biases among Indian Political Representatives. *Available at SSRN 4045630*.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes*, *101*(2), 127-151.

Bond, R., & Smith, P. B. (1996, January 1). Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task. *PSYCHOLOGICAL BULLETIN*, *119*(1), 111–137.

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLoS One*, *16*(6), e0253717.

Budescu, D. V., Rantilla, A. K., Yu, H. T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178-194.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*(1), 171-189

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta psychologica*, *104*(3), 371-398.

Buratti, S., Allwood, C. M., & Kleitman, S. (2013). First-and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition and learning*, 8(1), 79-102.

Chaiken, S. (1989). Heuristic and systematic information processing within and beyond the persuasion context. *Unintended thought*, 212-252.

Chen, Y. F. (2008). Herd behavior in purchasing books online. *Computers in Human Behavior*, *24*, 1977-1992.

Chiaburu, D.S., Stoverink, A.C., Li, N. and Zhang, X.-A. (2015), "Extraverts engage in more interpersonal citizenship when motivated to impression manage: getting along to get ahead?", *Journal of Management*, (in press).

Chiaburu, D. S., Cho, I., & Gardner, R. (2015b). Authenticity matters more than intelligence and personality in predicting metacognition. *Industrial and Commercial Training*, 47(7), 363-371.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621

Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, *25*(10), 1242-1253.

Cole, J., Watkins, C., & Kleine, D. (2016). Health advice from internet discussion forums: how bad is dangerous?. *Journal of medical Internet research*, *18*(1), e4.

Curtis, R. G., Windsor, T. D., & Soubelet, A. (2015). The relationship between Big-5 personality traits and cognitive ability in older adults–a review. *Aging, Neuropsychology, and Cognition*, *22*(1), 42-71.

Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The legal and ethical implications of using AI in hiring. *Harvard Business Review*, *25*.

DeCoster, J. (2012, June 19). *Stat-Help: Converting Effect Sizes*. Stat-Help.com - free statistics help over the internet. http://www.stat-help.com/spreadsheets/Converting%20effect%20sizes%202012-06-19.xls

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248-1299.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*(3), 629.

Devers, C., Steenbergh, T., Devers, E., Brighton, R., Ayers, B., Carmichael, E., ... & Hannum, P. (2012, October). Social conformity in online environments: A Comparison between text and video. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 1777-1781). Association for the Advancement of Computing in Education (AACE).

de Hooge, I. E., Verlegh, P. W. J., & Tzioti, S. C. (2014). Emotions in Advice Taking: The Roles of Agency and Valence: Emotions in Advice Taking. *Journal of Behavioral Decision Making*, *27*(3), 246–258. https://doi.org/10.1002/bdm.1801

De Wit, F. R. c., Scheepers, D., Ellemers, N., Sassenberg, K., & Scholl, A. (2017). Whether power holders construe their power as responsibility or opportunity influences their tendency to take advice from others. Journal of Organizational Behavior, 38(7), 923–949.

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological science*, *21*(6), 820-828.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one*, *18*(3), e0279720.

Duan, J., Gu, X., & Sun, L. (2016). The influence of explicit self-esteem, implicit self-esteem and the discrepancies between them on advice taking. *Acta Psychologica Sinica*, *48*(4), 371.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current directions in psychological science*, *12*(3), 83-87.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Vol. 44, pp. 247-296). Academic Press.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational behavior and human decision processes*, *105*(1), 98-121.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454-459.

Farjam, M., & Loxbo, K. (2023). Social conformity or attitude persistence? The bandwagon effect and the spiral of silence in a polarized context. *Journal of Elections, Public Opinion and Parties*, 1-21.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *TRENDS in Cognitive Sciences*, *8*(4).

Feher, A., & Vernon, P. A. (2021). Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality. *Personality and Individual Differences*, *169*, 110002.

Field, A.P. (2018) Discovering Statistics Using IBM SPSS Statistics. 5th Edition, Sage, Newbury Park.

Fieldhouse, E., Cutts, D., & Bailey, J. (2022). Who cares if you vote? Partisan pressure and social norms of voting. *Political Behavior*, *44*(3), 1297-1316.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in human neuroscience*, *8*, 443.

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying Associations between Conscientiousness and Executive Functioning: Mental Set Shifting, Not Prepotent Response Inhibition or Working Memory Updating. *Journal of Personality, 84*(3), 348–360. https://doi.org/10.1111/jopy.12163

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of consciousness*, *2017*(1), nix007.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25-42.

Galvin, S. J., Podd, J. V., Drga, V., and Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and error decisions. Psychon. Bull. Rev. 10, 843–876. doi: 10.3758/BF03196546

Gao, J., Yao, J., & Shao, Y. (2019, July). Towards reliable learning for high stakes applications. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 3614-3621).

Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. Applied Cognitive Psychology, 9, S55–S79.

Geisen, E. (2022, August 29). *Using attention checks in your surveys may harm data quality*. Qualtrics. https://www.qualtrics.com/blog/attention-checks-and-data-quality/

Gibbs, S., Moore, K., Steel, G., & McKinnon, A. (2017). The Dunning-Kruger Effect in a workplace computing setting. *Computers in Human Behavior*, *72*, 589-595.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*(1), 100-121.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21-35.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes*, *107*(2), 234–245. https://doi.org/10.1016/j.obhdp.2008.03.001

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology*, *102*(3), 497–512. https://doi.org/10.1037/a0026413

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: how emotions influence advice taking. *Journal of Applied Psychology*, *93*(5), 1165.

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational behavior and human decision processes*, *108*(2), 287-302.

Hamid, R. (2023, June 23). *Harvard Business School professor Francesca Gino accused of committing data fraud in at least four papers: News: The Harvard Crimson*. Harvard Business School Professor Francesca Gino Accused of Committing Data Fraud in at Least Four Papers. https://www.thecrimson.com/article/2023/6/23/alleged-data-fraud-gino/

Haran, U., & Shalvi, S. (2020). The implicit honesty premium: Why honest advice is more persuasive than highly informed advice. *Journal of Experimental Psychology: General*, *149*(4), 757–773. https://doi.org/10.1037/xge0000677

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making*, *17*(5), 333-348.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational behavior and human decision processes*, *70*(2), 117-133.

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational behavior and human decision processes*, *81*(2), 252-273.

Häusser, J. A., Leder, J., Ketturat, C., Dresler, M., & Faber, N. S. (2016). Sleep Deprivation and Advice Taking. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep24386

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*, *61*(3), 305-326

Hertz, N., & Wiese, E. (2018). Under pressure: Examining social conformity with computer and robot groups. *Human factors*, *60*(8), 1207-1218.

Himmelstein, M., Budescu, D., & Han, Y. (2023). The wisdom of timely crowds.

Hodges, B. H. (2014). Rethinking conformity and imitation: Divergence, convergence, and social understanding. *Frontiers in psychology*, *5*, 726.

Hofheinz, C., Germar, M., Schultze, T., Michalak, J., & Mojzisch, A. (2017). Are Depressed People More or Less Susceptible to Informational Social Influence? *Cognitive Therapy and Research*, *41*(5), 699–711. https://doi.org/10.1007/s10608-017-9848-7

Howard, J. (2019). Bandwagon effect and authority bias. In J. Howard (Ed.), Cognitive errors and diagnostic mistakes: A case-based guide to critical thinking in medicine (pp. 21–56). Springer.

Huang, M., Müller, C. L., & Gaynanova, I. (2021). latentcor: An R Package for estimating latent correlations from mixed data types. *arXiv preprint arXiv:2108.09180*.

Hughes, J. (2017). *paramtest: Run a Function Iteratively While Varying Parameters*. R package version 0.1.0, https://cran.r-project.org/web/packages/paramtest/paramtest.pdf

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: the Big Five revisited. *Journal of applied psychology*, *85*(6), 869.

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, *11*(4), 401–415. https://doi.org/10.1017/S193029750000382X

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate "micro" but not "macro" multicollinearity. *Behavior research methods*, *48*, 1308-1317.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual review of political science*, *22*, 129-146

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: Capturing decision tendencies in a fictitious medical test. *Metacognition and learning*, *9*, 25-49.

JASP Team (2022). JASP (Version 0.16.2) [Computer Software].

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2009). Can you be persuaded? individual differences in susceptibility to persuasion. In *Human-Computer Interaction–INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part I 12* (pp. 115-118). Springer Berlin Heidelberg.

Kastanakis, M. N., & Balabanis, G. (2012). Between the mass and the class: Antecedents of the "bandwagon" luxury consumption behavior. *Journal of Business Research*, *65*(10), 1399-1407.

Kausel, E. E., Culbertson, S. S., Leiva, P. I., Slaughter, J. E., & Jackson, A. T. (2015). Too arrogant for their own good? Why and when narcissists dismiss advice.

Organizational Behavior and Human Decision Processes, *131*, 33–50. https://doi.org/10.1016/j.obhdp.2015.07.006

Kelly, D., & Donaldson, D. (2016). Investigating the complexities of academic success: Personality constrains the effects of metacognition. *The Psychology of Education Review*, 40(2), 17-24.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81-93.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6).

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*(1), 140-151.

Krcmar, M., & Haberkorn, K. (2020). Mental Representations. The International Encyclopedia of Media Psychology, 1-17. https://doi.org/10.1002/9781119011071.iemp0191

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Krueger, J. I. (2003). Return of the ego--Self-referent information as a filter for social prediction: Comment on Karniol (2003).

Kuan, K. K., Zhong, Y., & Chau, P. Y. (2014). Informational and normative social influence in group-buying: Evidence from self-reported and EEG data. *Journal of Management Information Systems*, *30*(4), 151-178.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*(4), 241.

Lai, E. R. (2011). Metacognition: A literature review. *Always learning: Pearson research report*, *24*, 1-40.

Larsen, K. S. (1990). The Asch conformity experiment: Replication and transhistorical comparison. *Journal of Social Behavior and Personality*, *5*(4), 163.

Lazic, S.E. (2008). Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. *BMC physiology*, *8*(1), 1-7.

LeBeau, B. (2023). *simglm: Simulate Models Based on the Generalized Linear Model*. R package version 0.9.6, https://github.com/lebebr01/simglm.

Lewis, D. R. (2018). The perils of overconfidence: Why many consumers fail to seek advice when they really should. *Journal of Financial Services Marketing*, *23*, 104-111.

Lüdecke, D. (2023). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.15, https://CRAN.R-project.org/package=sjPlot.

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research*, *49*(2), 171-195. https://doi.org/10.1177/0093650220921321

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, *118*(23), e2019527118.

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, *10*(2), 3.

Marwade, A., Kumar, N., Mundada, S., & Aghav, J. (2017, September). Augmenting e-commerce product recommendations by analyzing customer personality. In *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)* (pp. 174-180). IEEE.

McFarlin, D. B., Baumeister, R. F., & Blascovich, J. (1984). On knowing when to quit: Task failure, self-esteem, advice, and nonproductive persistence. *Journal of Personality*, *52*(2), 138-155.

Meshi, D., Biele, G., Korn, C. W., & Heekeren, H. R. (2012). How Expert Advice Influences Decision Making. *PLOS ONE*, *7*(11), e49748. https://doi.org/10.1371/journal.pone.0049748

Modic D, Anderson R, Palomäki J (2018) We will make you like our research: The development of a susceptibility-to-persuasion scale. PLoS ONE 13(3): e0194119. https://doi.org/10.1371/journal.pone.0194119

Mohammadpour, S. I., & Nassiri, H. (2021). Aggressive driving: do driving overconfidence and aggressive thoughts behind the wheel, drive professionals off the road?. *Transportation research part F: traffic psychology and behaviour*, *79*, 170-184.

Moore, D. A., & Dev, A. S. (2017). Individual differences in overconfidence. *Encyclopedia of Personality and Individual Differences. Springer. Retrieved from http://osf. io/hzk6q*.

Moore, D. A., & Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, *11*(8), e12331.

Murata, A., Nakamura, T., & Karwowski, W. (2015). Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents. *Safety*, *1*(1), 44-58.

Neville, F. G., Templeton, A., Smith, J. R., & Louis, W. R. (2021). Social norms, social identities and the COVID-19 pandemic: Theory and recommendations. *Social and Personality Psychology Compass*, *15*(5), e12596.

Nguyen, N. T., Allen, L. C., & Fraccastoro, K. (2005). Personality predicts academic performance: Exploring the moderating role of gender. *Journal of Higher Education Policy and Management*, *27*(1), 105-117.

Olvera Astivia, O. L., Gadermann, A., & Guhn, M. (2019). The relationship between statistical power and predictor distribution in multilevel logistic regression: a simulation-based approach. *BMC medical research methodology*, *19*(1), 1-20.

Pavel, S. R., Robertson, M. F., & Harrison, B. T. (2012). The Dunning-Kruger effect and SIUC University's aviation participants. *Journal of Aviation Technology and Engineering*, *2*(1), 6.

Pimentel, R. S., Niewiadomska-Bugaj, M., & Wang, J. C. (2015). Association of zero-inflated continuous variables. *Statistics & Probability Letters*, *96*, 61-67.

Pimentel, R. S. (2009). *Kendall's Tau and Spearman's Rho for zero-inflated data*. Western Michigan University.

Pek, J., Wong, O., & Wong, A. C. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in psychology*, *9*, 2104.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality*, *88*(2), 185-200.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic bulletin & review*, *24*, 1774-1784.

Perfumi, S. C., Bagnoli, F., Caudek, C., & Guazzini, A. (2019). Deindividuation effects on normative and informational social influence within computer-mediated-communication. *Computers in human behavior*, *92*, 230-237.

Pescetelli, N., Hauperich, A. K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. *Cognition*, *215*, 104810.

Petro, G. (2017, November 21). *Wisdom of the crowd: From TV show to Retail Business Model*. Forbes. Retrieved March 31, 2023, from https://www.forbes.com/sites/gregpetro/2017/11/21/wisdom-of-the-crowd-from-tv-show-to-retail-business-model/

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic. Journal of Behavioral Decision Making, *17*, 39–57.

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, *130*, 26–43. https://doi.org/10.1016/j.obhdp.2015.05.004

Redelmeier, D. A., & Scales, D. C. (2015). Missing the diagnosis of brain death as a self-erasing error. *American Journal of Respiratory and Critical Care Medicine*, *192*(3), 280-282.

Rees, L., Rothman, N. B., Lehavy, R., & Sanchez-Burks, J. (2013). The ambivalent mind can be a wise mind: Emotional ambivalence increases judgment accuracy. *Journal of Experimental Social Psychology*, *49*(3), 360–367. https://doi.org/10.1016/j.jesp.2012.12.017

Reiling, A. D. (2020). Courts and artificial intelligence. In *IJCA* (Vol. 11, p. 1).

Ripberger, J. T., Silva, C. L., Jenkins-Smith, H. C., Carlson, D. E., James, M., & Herron, K. G. (2015). False alarms and missed events: The impact and origins of perceived inaccuracy in tornado warning systems. *Risk analysis*, *35*(1), 44-56.

Rosander, M., & Eriksson, O. (2012). Conformity on the Internet – The role of task difficulty and gender differences. *Computers in Human Behavior*, *28*(5), 1587–1595. https://doi-org.libproxy.unm.edu/10.1016/j.chb.2012.03.023

Ross, S. M. (2022). *Simulation*. Academic Press.

Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, *121*(2), 246-255.

Salomons, N., Sebo, S. S., Qin, M., & Scassellati, B. (2021). A minority of one against a majority of robots: Robots cause normative and informational conformity. *ACM Transactions on Human-Robot Interaction (THRI)*, *10*(2), 1-22.

Sanchez, D.N., (2021). Are the Low-Skilled the Only Ones Blind to their Errors? Unpacking the Dunning-Kruger Effect Using Trial-by-Trial Judgments of Performance. Unpublished manuscript.

Sanchez, D.N., Benson, M., Ruthruff, E. (2023). Are Errors Invisible only to Poor Performers? Unpacking the Dunning-Kruger Effect Using Trial-by-Trial Confidence in Performance. Manuscript submitted for publication.

Sanchez, D.N., & Speed, A. (2020). *The Impact of Individual Traits on Domain Task Performance: Exploring the Dunning-Kruger Effect* (No. SAND-2020-12152). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Sarkar, S., Aleali, A., Shakarian, P., Armenta, M., Sanchez, D., & Lakkaraju, K. (2019, August). Impact of social influence on adoption behavior: An online controlled experimental evaluation. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 226-233).

Sauer, S. (2017, January 24). *Convert logit to probability*. Sebastian Sauer Stats Blog. https://sebastiansauer.github.io/convert_logit2prob/

Saylik, R., Szameitat, A. J., & Cheeta, S. (2018). Neuroticism related differences in working memory tasks. *PloS one*, *13*(12), e0208248.

Scheiber, N. (2023, June 24). *Harvard scholar who studies honesty is accused of fabricating findings*. The New York Times. https://www.nytimes.com/2023/06/24/business/economy/francesca-gino-harvard-dishonesty.html

Scheunemann, J., Gawęda, Ł., Reininger, K.-M., Jelinek, L., Hildebrandt, H., & Moritz, S. (2020). Advice weighting as a novel measure for belief flexibility in people with psychotic-like experiences. *Schizophrenia Research*, *216*, 129–137. https://doi.org/10.1016/j.schres.2019.12.016

Schultze, T., & Loschelder, D. D. (2021). How numeric advice precision affects advice taking. *Journal of Behavioral Decision Making*, *34*(3), 303–310. https://doi.org/10.1002/bdm.2211

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the Inability to Ignore Useless Advice: A Case for Anchoring in the Judge-Advisor-System. *Experimental Psychology*, *64*(3), 170–183. https://doi.org/10.1027/1618-3169/a000361

See, K. E., Morrison, E. W., Rothman, N. B., & Soll, J. B. (2011). The detrimental effects of power on confidence, advice taking, and accuracy. *Organizational behavior and human decision processes*, *116*(2), 272-285.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational behavior and human decision processes*, *62*(2), 159-174.

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational behavior and human decision processes*, *84*(2), 288-307.

Sniezek, J. A., Schrah, G. E., & Dalal, R. S. (2004). Improving judgement with prepaid expert advice. *Journal of Behavioral Decision Making*, *17*(3), 173–190. https://doi.org/10.1002/bdm.468

Soll, J. B., & Larrick, R. (1999, November). The 80/20 rule and revision of judgment in light of another's opinion: why do we believe in ourselves so much? Paper presented at the annual meeting of the Society for Judgment and Decision Making, Los Angeles, CA.

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.

Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: a simplified three-step procedure using Stata, R, mplus, and SPSS. *International Review of Social Psychology*, *30*, 203-218.

Sowden, S., Koletsi, S., Lymberopoulos, E., Militaru, E., Catmur, C., & Bird, G. (2018). Quantifying compliance and acceptance through public and private social conformity. *Consciousness and cognition*, *65*, 359-367.

Sridhar, S., & Srinivasan, R. (2012). Social Influence Effects in Online Product Ratings. *Journal of Marketing*, *76*(5), 70–88. https://doi.org/10.1509/jm.10.0377

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Stankov, L., Kleitman, S., & Jackson, S. A. (2015). Measures of the trait of confidence. *Measures of personality and social psychological constructs*, 158-189.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, *53*(1), 44-49.

Sullivan, P. J., Ragogna, M., & Dithurbide, L. (2019). An investigation into the Dunning–Kruger effect in sport coaching. *International Journal of Sport and Exercise Psychology*, *17*(6), 591-599.

Sukenik, S., Reizer, A., & Koslovsky, M. (2018). Direct and indirect effects of Agreeableness on overconfidence. *Journal of Individual Differences*.

Sutherland, S. C., Harteveld, C., & Young, M. E. (2016). Effects of the advisor and environment on requesting and complying with automated advice. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *6*(4), 1-36.

Tauni, M.Z., Ali, F., Yousaf, S., Shaikh, H.A., & Zia-Ur-Rehman, R. A. O. (2020). The moderating role of advisor big five personality on the association between financial advice and investor trading: evidence from the chinese futures market. *Psychologia*, *62*, 289-311.

The jamovi project (2023). *jamovi* (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Thomson, K. S., & Oppenheimer, D. M. (2016). Cognitive Reflection Test-2. *Judgment and Decision Making*.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & reasoning*, *20*(2), 147-168.

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational behavior and human decision processes*, *117*(1), 53-65.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124-1131.

Tzini, K., & Jain, K. (2018). The Role of Anticipated Regret in Advice Taking. *Journal of Behavioral Decision Making*, *31*(1), 74–86. https://doi.org/10.1002/bdm.2048

Van Horn, L. (2023). Converting Logistic Multi-Level Model Logit Outputs to Probabilities (Version 1.0) [R Source Code].

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British journal of social psychology*, *44*(3), 443-46

Vuorre, M., & Metcalfe, J. (2021). Measures of relative metacognitive accuracy are confounded with task performance in tasks that permit guessing. *Metacognition and Learning*, 1-23.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192-196.

Walters, D. J., Fernbach, P. M., Fox, C. R., & Sloman, S. A. (2017). Known unknowns: A critical determinant of confidence and calibration. *Management Science*, *63*(12), 4298-4307.

Wang, S., Chu, T. H., & Huang, G. (2023). Do Bandwagon Cues Affect Credibility Perceptions? A Meta-Analysis of the Experimental Evidence. *Communication Research*, 00936502221124395.

Wang, X., & Du, X. (2018). Why Does Advice Discounting Occur? The Combined Roles of Confidence and Trust. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02381

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143* (5), 2020-2045.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

Wijenayake, S., van Berkel, N., Kostakos, V., & Goncalves, J. (2020). Impact of contextual and personal determinants on online social conformity. *Computers in Human Behavior*, *108*, 106302.

Wijenayake, S., Hu, J., Kostakos, V., & Goncalves, J. (2021). Quantifying the effects of age-related stereotypes on online social conformity. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part IV 18* (pp. 451-475). Springer International Publishing.

Wilt, J. and Revelle, W. (2009), "Extraversion", in Leary, M.R. and Hoyle, R.H. (Eds), Handbook of Individual Differences in Social Behavior, The Guilford Press, New York, NY, pp. 27-45.

Winne, P. H. (1996). A Metacognitive View of Individual Differences in Self-Regulated Learning. LEARNING AND INDIVIDUAL DIFFERENCES, (4), 327.

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior research methods*, *41*, 257-267.

Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. *Progress in molecular biology and translational science*, *171*, 309-491.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational behavior and human decision processes*, *93*(1), 1-13.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, *83*(2), 260-281.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104-120.

Zell, E., Strickhouser, J. E., Sedikides, C., & Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological bulletin*, *146*(2), 118.

Zell, E., & Lesick, T. L. (2022). Big five personality traits and performance: A quantitative synthesis of 50+ meta-analyses. *Journal of personality, 90*(4), 559-573.

Zhang, T., Harrington, K. B., & Sherf, E. N. (2022). The errors of experts: When expertise hinders effective provision and seeking of advice and feedback. *Current Opinion in Psychology*, *43*, 91-95.

Zhang, K. Z., Zhao, S. J., Cheung, C. M., & Lee, M. K. (2014). Examining the influence of online reviews on consumers' decision-making: A heuristic–systematic model. *Decision Support Systems*, *67*, 78-89.