Debias Human Judgment with a Common Decision Bias:
An Experimental Study of the Anchoring Effect in Hiring Decision-Making

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Siwen Shu

IN PARTIAL FULTILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Nathan R. Kuncel, Adviser

May 2018

Acknowledgements

I would like to extend thanks to everyone who has supported me throughout my academic career. A special thanks to my adviser, Dr. Nathan Kuncel, for his guidance and encouragement over the years. I would also like to thank the members of my dissertation committee, Dr. Paul Sackett, Dr. Aaron Schmidt, and Dr. Pri Shah, for their valuable feedback throughout the process.

In addition, I would like to thank other faculty of the I/O psychology program at the University of Minnesota, Dr. John Campbell and Dr. Deniz Ones, for providing great opportunities to enrich my knowledge and skills in the field. I would also like to thank my fellow students in the I/O psychology program for their academic and emotional support along the way.

Finally, I would like to thank my family and friends for their companionship throughout my life.

Abstract

Hiring decision-making is heavily based on intuition-based approaches (e.g., expert judgment, group consensus meetings) despite extensive research supporting the use of analytical approaches (e.g., algorithms, statistical equations). Both individual level and organizational level factors contribute to practitioners' resistance to using statistical methods. Given that the reliance on human intuition is likely to persist, it is important to identify interventions that improve the accuracy of human judgment when making hiring decisions. The purpose of this study is to examine the effectiveness of eliciting a favorable anchoring effect to improve hiring decisions. Although the anchoring effect is a common decision bias, it may increase the accuracy of human predictions, which are aligned with the anchors, if the anchors are valid predictions of the outcome. Specifically, four questions are examined: 1) Does providing decision makers with external predictions (i.e., anchors) affect the decision policy and the accuracy of human judgment? 2) Do features of the anchors (i.e., source and precision) affect the degree of anchoring and the accuracy of human judgment? 3) Do individual differences predict anchoring susceptibility? and 4) Do individual differences interact with anchor features when predicting the degree of anchoring? Results showed that providing decision makers with external predictions as a decision aid induced favorable anchoring. Furthermore, it increased the consistency and accuracy of human judgment as well as influenced the weighting policy of decision makers. The more precise the anchors were, the stronger the anchoring effect was. Knowing the source of the anchors did not affect the degree to which decision makers adopted the anchors. Different effects of anchor features were observed when predicting accuracy: when anchors were presented at a precise level, individuals who were informed that the anchors were provided by an expert showed lower accuracy than those who were informed that the anchors were obtained from a statistical formula. In fact, there was only a weak linear association between degree of anchoring and accuracy because individuals could outperform the suboptimal anchors (i.e., unit-weighted composites of the predictors) presented to them. Exploratory analyses revealed that stronger anchoring led to higher accuracy only when individuals did not develop a better weighting schema than unit weighting. When individuals were able to apply a weighting schema that was more accurate than unit weighting, weaker anchoring implied reasonable adjustment and thus higher accuracy. Prior experience with employee selection predicted anchoring susceptibility such that individuals with more prior experience showed weaker anchoring. Cognitive ability interacted with the source of anchors when predicting anchoring such that higher cognitive ability was related to weaker anchoring only when individuals were informed that the anchors were obtained from a statistical formula.

Table of Contents

List of Tables

List of Figures

Literature Review

Despite the extensive research that mechanical (i.e., statistical) combinations of predictors are consistently equal to or superior to human judgment (e.g., Grove, Zald, Lebow, Snitz, & Nelson, 2000), most decisions made in work settings are heavily based on human intuition. Various explanations have been offered to explain the resistance to mechanical methods[1]. Besides apparent factors, such as lacking knowledge and education of statistical methods (Vrieze & Grove, 2009) and belief in human intuition (Highhouse, 2008), there are contextual reasons for this resistance. In an employment setting, selection is fundamentally a sociopolitical process (Cleveland & Murphy, 1992). Organizational factors, such as politics and culture, also promote the use of human judgment while discouraging the use of purely statistical methods (Johns, 1993; Muchinsky, 2004; Terpstra & Rozell, 1997).

Given that relying solely on statistical methods, especially in high-stakes settings, is unlikely as a practical matter, it is crucial to understand what interventions may help decision makers effectively integrate data and make more accurate judgments in the context of hiring decision-making. Ideally, such an approach would preserve the predictive power of an algorithmic method while remaining acceptable to end users.

---

[1] Sawyer (1966) distinguished between mode of data collection and mode of data combination. This study is focused on comparisons of clinical and mechanical methods of data combination. Mechanical methods refer to any processes involving integrating data with consistent rules (e.g., algorithms, statistical equations, models of human judges, etc.), regardless of whether the data is collected clinically (e.g., interviews) or statistically (e.g., psychometric tests).

Evidence suggests that interventions such as task-related feedback and training on cue weighting strategies affect human judgment but improvements in accuracy are typically small (Harding, 2004; Karelaia & Hogarth, 2008). Practically speaking, immediate feedback is rarely available to decision makers after making hiring decisions since there is often a time lag between when the decision is made and when the criterion can be measured. In some cases, the decision maker will have no contact with the employee after the decision is made. Similarly, although research suggests it can be helpful to provide judges with training on domain-specific knowledge, such as how predictors are associated with the criterion or how predictors should be integrated statistically, the practical challenge is that developing and implementing such training can be costly and time-consuming.

A major obstacle to improving hiring decision-making is that human judges are subject to a range of decision biases. According to the theory of bounded rationality, human judgment is rational only within the limitations imposed by the amount of information they have and the cognitive capacities of their minds (Simon, 1955). To cope with these limitations when making decisions, instead of using more complex rules and algorithms, individuals tend to use fast heuristics to reduce the complexity of decision processes. Heuristics refer to methods that allow individuals to reach satisfactory solutions with modest amounts of processing (Simon, 1990). Examples include representativeness, availability, and anchoring-and-adjustment heuristics (Tversky & Kahneman, 1974). Although heuristics allow individuals to make decisions in a less effortful manner and are sometimes useful, they can lead to systematic biases. For instance, the anchoring-and-adjustment heuristic refers to the tendency of decision makers to make judgments biased toward an initially presented value (Tversky & Kahneman, 1974). Although individuals do adjust from the starting value, the adjustments are typically insufficient. If the initially presented value is biased or irrelevant to the target judgment, the human judgment fixated on the anchor will also be biased.

The anchoring heuristic is a ubiquitous phenomenon in human judgment. Efforts have been devoted to mitigating the anchoring effect (e.g., Galinsky & Mussweiler, 2001; Mussweiler, Strack, & Pfeiffer, 2000) but the effect remains robust. If we cannot effectively prevent it, can we leverage the anchoring bias to improve human judgment? If human judgment can be substantially influenced by an externally provided anchor, one can argue that providing the prediction from an external source to the human judge prior to making decisions will also induce the anchoring effect and encourage the human judge to align his or her judgment toward the external prediction. If the externally provided predictions have a reasonable level of validity for predicting the outcome of interest, human judgments, when aligned with external predictions, should also be predictive of the outcome. In other words, the accuracy of human judgment should be improved due to anchoring when the anchor is highly valid.

Although the anchoring effect is robust, recent research suggests the magnitude of anchoring varies depending on the features of anchors (Dowd, Petrocelli, & Wood, 2014; Janiszewski & Uy, 2008) and there is a considerable amount of individual differences in the susceptibility to anchoring (Furnham, Boo, & McClelland, 2012; Wilson, Houston, Etling, & Brekke, 1996). Therefore, understanding what factors boost the anchoring effect is also crucial for eliciting a favorable anchoring to improve hiring decision-making.

The objective of this study is to examine the effectiveness of eliciting a favorable anchoring effect to improve hiring decision-making. More specifically, four research questions are examined: 1) Does providing human judges with predictions from an external source, which serve as anchors upon which one can adjust their judgments, affect their decision policy and judgment accuracy? 2) Do features of the anchor influence the magnitude of anchoring and the accuracy of human judgment? 3) Do individual differences predict susceptibility to anchoring? 4) How do individual differences interact with features of the anchor when predicting the magnitude of anchoring?

I next review the literature to integrate the available information in one place and help frame the discussion of the current study. First, the anchoring research is introduced to explain the underlying mechanisms of anchoring and why providing anchors to human judges could potentially improve decision-making. Next, I discuss the extensive research comparing clinical and mechanical methods to illustrate how the specific nature of an anchor affects the effectiveness of its use to improve human judgment. I then review the existing literature on the clinical synthesis, a data combination method in decision-making which shares the fundamental principles with the intervention proposed in this study, to demonstrate the feasibility of using anchoring to improve human judgment. Lastly, I discuss how individual differences play important roles in the susceptibility to anchoring and the effectiveness of anchoring for improving hiring decision-making.

Anchoring Effect

The anchoring effect was first introduced by Tversky and Kahneman (1974) as the anchoring-and-adjustment heuristics, and has since then been replicated in various domains, such as general knowledge (Strack & Mussweiler, 1997) and probability estimates (Chapman & Johnson, 1999). The effects are not limited to laboratory settings, but have been demonstrated with real-world judgments, such as negotiation (Galinsky & Mussweiler, 2001), judicial decisions (Englich & Soder, 2009), and forecasting (Critcher & Gilovich, 2008).

In the classic study by Tversky and Kahneman (1974), participants were asked to estimate the percentage of African countries in the United Nations. Before actually making the judgment, participants were asked to determine whether an arbitrary starting value (a value between 0 and 100 determined by spinning a wheel of fortune) was higher or lower than the percentage to be estimated. They found that participants' estimates were significantly biased towards the arbitrary starting value. Most anchoring studies have followed this two-step procedure where participants first make a comparative judgment, comparing the target value with the anchor value, and then make an absolute judgment of the target value. However, it has been demonstrated that the anchoring effect could occur without the process of explicitly comparing the anchor with the target, but the anchor has to be very salient by extensive processing (Wilson et al., 1996). For instance, Wilson et al. demonstrated that a quick appraisal of an arbitrary anchor (e.g., counting its number of digits) did not cause anchoring, while paying more attention to the number (e.g., comparing the anchor with a closely related number) was sufficient to induce anchoring although a direct comparison to the target was absent. However, it was not clear how much attention to the anchor would be sufficient.

The robustness of the anchoring effect in very different domains raises the question, "Why are individuals susceptible to the influence of anchors?". Multiple theories have been proposed to account for this robust effect but consensus has not been reached regarding what mechanisms contribute to anchoring. To fully understand how the anchoring effect can be applied to hiring decision-making, it is important to first review those theoretical accounts. I then discuss how the features and presentations of anchors may influence the magnitude of anchoring and their implications for eliciting a favorable anchoring effect to improve hiring decision-making.

Theories of Anchoring Effects

Insufficient Adjustment

One of the early explanations is that anchoring results from an effortful process of adjustment: people adjust their estimate from the anchor value toward the range of plausible values, but this adjustment is typically insufficient and therefore the final estimate is biased toward the anchor (Tversky & Kahneman, 1974). However, this mechanism is not applicable to many situations. For instance, the process of adjustment cannot explain why anchoring effects occur with both plausible and implausible anchors. Individuals adjust when they think the anchor value falls outside of the plausible range of values. With plausible anchors, there is no reason to adjust in the first place as the anchor is already within the range of possible answers. As a result, the adjustment process is only applicable when the anchor is more extreme than the boundary of plausible values (Strack & Mussweiler, 1997). Moreover, the finding that anchoring effects can be induced with subliminal anchors indicates that the conscious process of adjustment is not necessary for anchoring (Mussweiler & Englich, 2005).

Selective Availability

The second theory of anchoring effects, which has received extensive empirical support, posits that anchoring occurs because the anchor activates information of the target that is consistent with the anchor (Mussweiler & Strack, 1999; Strack & Mussweiler, 1997). More specifically, when presented with an anchor, individuals engage in a hypothesis testing process where they test the hypothesis that the anchor is the correct answer for the target judgment (Chapman & Johnson, 1994). This process of confirmatory search increases the availability of similar features between the anchor and the target and reduces the availability of features of the target that differ from the anchor (Chapman & Johnson, 1999). Subsequently, when individuals make a final judgment, they rely heavily on information that is most accessible (Higgins, 1996). Therefore, the judgment is biased toward the anchor as the most accessible information is the anchor-consistent information of the target activated by the anchor.

According to the selective availability explanation, anchoring should diminish if the information activated by the comparative judgment is irrelevant to the target of the absolute judgment. For instance, Strack and Mussweiler (1997) found that presenting anchors of the width of the Brandenburg Gate had a larger effect on judgments of its width but only a small effect on judgments of its height. Moreover, anchoring should not occur if the anchor activates information that overlaps with what would have been activated by the judgment itself without the anchor. In support of this, Bahník and Strack (2016) found that anchoring effect diminished when the comparative judgment concerned a general category and the target of the absolute judgment was a subset of the general category that was activated by the anchor. In one of the experiments, participants were asked to compare the average annual temperature in New York City to either a low (–4 °F) or a high (102 °F) anchor. They found that for participants who received the high anchor value (102 °F) anchoring occurred for the judgment of the average winter (average prediction in the high anchor group was 35.1 °F as compared to 29.5 °F in the control group) but not summer temperature (average prediction in the high anchor group was 81.5 °F as compared to 82.7 °F in the control group). The high anchor value activated information about summer, which overlapped with information that would have been activated when making the absolute judgment of the average summer temperature without the anchor and therefore did not induce anchoring. On the other hand, the information about summer activated by the high anchor did not overlap with information activated when making the absolute judgment of the average winter temperature and therefore induced anchoring. Similarly, it was found that participants who were presented with the low anchor (–4 °F) showed an anchoring effect only when judging the average summer temperature (average prediction in the low anchor group was 78.4 °F as compared to 82.7 °F in the control group) but not winter temperature (average prediction in the low anchor group was 30.4 °F as compared to 29.5 °F in the control group). The low anchor value activated information about winter that overlapped with information that would have been activated when predicting the average winter temperature without the anchor and therefore did not induce anchoring. On the contrary, predicting the average summer temperature activated information about summer while the low anchor activated information about winter. Because the anchor and the absolute judgment did not activate the same information, anchoring was induced by the low anchor.

However, the selective availability account has been questioned by recent findings that anchoring effects can occur even when the target of the comparative judgment is not the same as the target of the subsequent absolute judgment. For instance, similar anchoring effects were observed when judging the price of a camera regardless of whether the preceding comparative judgment was about a GPS device or a camera (Mochon & Frederick, 2013). This finding suggests that selective availability is, at least, not the only mechanism underlying anchoring effects.

Attitude Change

A more recent perspective suggests that anchors serve multiple roles when inducing anchoring effects and anchoring can be a result of thoughtful, high-elaboration processes, or non-thoughtful, low-elaboration processes (Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010). High-elaboration processes, such as confirmatory search and selective accessibility, do not have to occur for anchoring. When individuals lack motivation or ability to elaborate, anchoring occurs due to non-thoughtful processes. For example, individuals may consider the anchor as a hint to the correct answer without consciously evaluating whether the anchor is plausible or relevant when lacking motivation or ability (Schwarz, 1994). The notion of thoughtful and non-thoughtful processes of anchoring has received some empirical support (Petty & Cacioppo, 1986; Wegener, Blankenship, Petty, & Detweiler-Bedell, 2009) and this attitude change perspective may serve as a promising account of anchoring effects, especially when the anchor is extreme and implausible.

To conclude, different theories have been proposed to explain the underlying process of anchoring although none of them seem to be ubiquitously applicable. Given that anchoring effects have been demonstrated with various paradigms in various domains, it is understandable that there is no single theory that explains the effect unconditionally. More research should investigate under what circumstances each of these theories is a better account of anchoring effects. Regardless of theoretical accounts, the robustness of anchoring effects holds in a range of fields, but this does not preclude the effect from being moderated by various factors. Prior literature suggests how anchors are presented affects the magnitude of anchoring. Two features that are most relevant to the current study are anchor precision and anchor relevance.

Anchor Precision

In a study of general knowledge judgment, Janiszewski and Uy (2008) found that anchor precision influenced the magnitude of anchoring such that a precise anchor led to less adjustment and thus a stronger anchoring effect than a rounded anchor. For instance, when estimating the price of a product (e.g., a plasma TV), participants showed a stronger anchoring effect when the anchor was presented as a precise number ($4,998) instead of a rounded number ($5,000). They demonstrated that this effect was not due to the fact that a precise anchor might result in a narrower range of plausible values than a rounded anchor, or a precise anchor might be perceived as more accurate or valid than a rounded anchor. In fact, anchor precision impacted the resolution of the underlying subjective response scale: a precise anchor results in a finer-grained subjective scale where individuals adjust away from the anchor in smaller units, while a rounded anchor results in a more coarse-grained scale where adjustment occurs with larger units.

However, research suggests that precision only influences the magnitude of anchoring effects when the anchor is perceived as informative or relevant to the target judgment (Zhang & Schwarz, 2013).

The finding that precise anchors lead to stronger anchoring was also supported by studies of price negotiation, where offer recipients tended to make greater counteroffer adjustments, and therefore showed weaker anchoring effects, to round rather than precise offers (Loschelder, Stuppi,& Trötschel, 2014; Mason, Lee, Wiley, & Ames, 2013). However, a recent study suggests precision could backfire with experts. Loschelder, Friese, Schaerer, and Galinsky (2016) showed that in the context of price negotiation, a more precise first offer led to stronger anchoring effects for amateurs but an inverted-U-shaped effects for experts. When the recipients were experts, increased precision first boosted anchoring but eventually reduced the anchoring effect. One explanation is that a precise offer implies more knowledge and confidence than a round offer (Jerez-Fernandez, Angulo, & Oppenheimer, 2014; Mason et al., 2013). However, when there is too much precision, experts may perceive it as reflecting a lack of competence.

The implication for the proposed intervention is that when eliciting a favorable anchoring effect to improve hiring decision-making, presenting the externally provided predictions at a more precise level should induce a stronger anchoring effect and therefore a greater improvement in judgment accuracy. Thus, the externally provided predictions were presented at different levels of precision in the present study to examine how precision affects the magnitude of anchoring and the accuracy of human judgments.

Anchor Relevance

A large body of research has demonstrated that anchoring occurs even when the anchor is completely irrelevant and random (Critcher & Gilovich, 2008; Tversky & Kahneman, 1974). However, recent research suggests that anchor relevance may moderate anchoring effects such that relevant anchors lead to stronger anchoring effects than irrelevant anchors. Earlier studies have showed mixed findings of the effect of relevance on anchoring, presumably due to small sample sizes and reliance on participants' subjective ratings of relevance (e.g., Chapman & Johnson, 1999; Englich & Mussweiler, 2001).

In a recent study by Glöckner and Englich (2015), participants were instructed to make sentencing decisions based on case reports and sentencing guidelines. To understand the effect of relevance, participants were randomly assigned to receive no anchor, an irrelevant anchor (judgment of an unrelated case from a layperson), or a relevant anchor (judgment from an expert). Glöckner and Englich found that anchoring effects were stronger when anchors were relevant rather than irrelevant. The finding that relevance matters for anchoring is in line with the insufficient adjustment perspective as individuals are less likely to adjust away from the anchor value if they consider the anchor as relevant. In addition, it is also consistent with the attitude change perspective in that if individuals consider the anchor as relevant they should be more likely to consider the anchor as a hint of the correct answer for the target judgment (Wegener et al., 2010).

In addition to legal judgments, the impact of anchor relevance has been found in other domains. For instance, when estimating the population of Chicago, IL, participants placed greater weight on the externally provided anchor when it was from a highly credible source than a less credible source (Dowd et al., 2014). In another setting where participants estimated the maximum price they would pay for a product, Bavoľár (2017) found that the anchoring effect was stronger when the anchor was presented as the price paid by a hypothetical person (relevant) than when the anchor was from a random mathematical task (irrelevant).

If the relevance of anchors matters, one can argue that providing predictions from an external source, which function as relevant anchors, to the human judge as a decision aid will also induce the anchoring effect. In fact, the idea of providing external predictions to assist human judgment is not new. A version of it was proposed by Sawyer (1966) who proposed the clinical synthesis method as an alternative to both the clinical method and the mechanical method when integrating data to reach an overall prediction. The clinical synthesis refers to the process of providing a statistically combined prediction to the human judge and the judge could then integrate the statistical prediction with the rest of the data. The statistical prediction is obtained by applying any consistent rules (e.g., algorithms, statistical equations, models of human judges, etc.) to integrate the data including all relevant predictors and it is essentially an anchor upon which the judge can make reasonable adjustments. Here the objective is not that the anchored judgment is more accurate than a mechanical composite but it is more accurate than what the judge would have produced without the anchor.

Although the concept was proposed decades ago, only a few studies have examined the predictive power of the clinical synthesis above and beyond purely clinical and mechanical methods with mixed results (e.g., Ægisdóttir et al., 2006; Leli & Filskov, 1984; Melton, 1952). One explanation for the scarcity of research is decision makers' preference for using holistic approaches (i.e., human intuition) instead of analytical approaches (i.e., algorithms or statistical models) in employment decisions (Highhouse, 2008; Nolan, 2012). As a result, decision makers are less likely to adopt a decision aid if it stems from a purely statistical source (e.g., statistically combined predictions) instead of a more intuition-based source (e.g., expert predictions). The present study examines how the source of anchors affects the extent to which decision makers accept externally provided anchors (i.e., the magnitude of anchoring). Although the external predictions were obtained based on a statistical model, information about the external source was manipulated to vary along a holistic-analytical continuum (i.e., a statistical model, expert judgment, or a combination of a statistical method and expert judgment).

Given practitioners' preference for intuition-based rather than analytical approaches, I propose that judges will perceive expert judgment as the most credible and acceptable source and therefore show the greatest anchoring effect when being informed that the external predictions are based on expert judgment. Whether the source of anchors affects the degree of anchoring also provides a test of the three different theories of anchoring. According to the insufficient adjustment account, the source of anchors matters as the more credible individuals perceive a source is the less likely they would adjust from the anchor and therefore show stronger anchoring. Similarly, according to the attitude change account, the source of anchors matters because the more credible individuals perceive a source is, the more likely they would believe the anchor is a hint to the correct answer and therefore show stronger anchoring. On the other hand, the selective availability account would predict that the source of anchors does not necessarily affect the degree of anchoring. When one compares the target with the anchor value, the accessibility of anchor-consistent information about the target is increased regardless of whether the anchor is informative (Englich & Mussweiler, 2001).

To fully understand how the tension between holistic and analytical approaches affects employment decisions and the implications for eliciting a favorable anchoring effect to improve hiring decisions, I next review the extensive research on the clinical and mechanical prediction debate. Prior literature on the clinical synthesis will also be discussed to demonstrate the feasibility of using anchoring to improve human judgment.

Clinical versus Mechanical Prediction

The debate over clinical versus mechanical prediction was first broadly introduced in Paul Meehl's (1954) seminal work, *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. The clinical prediction (i.e., expert judgment, holistic method, intuition-based approach, etc.) refers to any processes relying on human judgment and intuition to integrate data in order to make an overall prediction. In contrast, the mechanical prediction (i.e., statistical, actuarial, or algorithmic prediction, etc.) refers to any processes using algorithms or statistical formulas to integrate data. Meehl (1954) reviewed 20 studies in the literature and showed that the mechanical method was equal to or superior to the clinical method in every case. Since then, the robust superiority of mechanical methods over clinical methods has been demonstrated across a range of fields, such as physical and mental health diagnoses (Grove & Meehl, 1996; Grove et al., 2000), hiring and admission decisions (Highhouse & Kostek, 2013; Kuncel, Klieger, Connelly, & Ones, 2013).

Focusing on predicting human health and behavior, Grove et al. (2000) demonstrated that mechanical predictions were consistently as accurate as or more accurate than clinical predictions across judgment tasks, types of judges, and types of data being combined. However, most studies included in this meta-analysis were health-related and only a few studies made predictions of work outcomes. Borman (1982) found that when predicting training performance, statistical composites of assessment ratings were slightly more valid than the consensus judgment by multiple assessors. Similarly, a primary study by Ganzach, Kluger, and Klayman (2000) illustrated that the mechanical combination of information collected in a structured interview outperformed clinical judgment. A recent meta-analysis examined this issue in the context of job selections and academic admissions (Kuncel et al., 2013). In the context of selection and admissions, mechanical data combination methods refer to reaching a selection decision by statistical models, such as multiple regression, unit-weighting, differential weighting, and bootstrapped weighting of the predictors (e.g., scores collected via tests, interviews, resumes etc.). Clinical combinations refer to reaching a selection decision by one or multiple human assessors who have access to the data of the predictors. Consistent with previous research, Kuncel et al. found that mechanical combinations yielded larger validities than clinical combinations for most criteria and equal validities for a few criteria. There is clear evidence that mechanical approaches in general have stronger correlations with standard validation criteria in the context of hiring decision-making.

One major explanation for the superiority of mechanical methods is human cognitive limitations. Compared with statistical models, human judges are much more inconsistent in applying weighting schemas across judgments (Grove & Meehl, 1996). They also tend to overemphasize salient cues that are not necessarily predictive of the outcome (Kuncel et al., 2013) and believe in illusory relationships (Chapman &

Chapman, 1969). Both result in less optimal weights than statistically modeled weights. As discussed earlier, because of limited cognitive capacities, decision makers often engage in fast heuristics instead of comprehensive analysis when making decisions. Although heuristics often produce "good enough" decisions, they lead to systematic biases especially in complex scenarios. For instance, the representativeness heuristic refers to the tendency of estimating the probability of one event based on whether the event is representative of another event. An example provided by Tversky and Kahneman (1974) is that "consider an individual, Mr. X, who has been described as 'meticulous, introverted, meek, solemn', and the following set of occupational roles: farmer, salesman, pilot, librarian, physician. How do people evaluate the likelihood that Mr. X is engaged in each of these occupations, and how do they order the occupations in terms of likelihood?" (p. 3). According to the representativeness heuristic, people will rank order the likelihood of each occupation by the extent to which Mr. X is representative of the stereotypes associated with each occupation. Because the personality descriptions of Mr. X are representative of the stereotype of librarians, people are more likely to predict Mr. X as a librarian due to representativeness bias. However, representativeness or similarity should not be confounded with probability. When estimating the probability that Mr. X is a librarian versus a farmer, the fact that the base rate of farmers is higher than the base rate of librarians should be considered even though Mr. X is more similar to the stereotypes of librarians than farmers. Ignoring the base rate of outcomes would lead to biased estimates. In the context of hiring decisions, if decision makers possess stereotypes of what an ideal candidate should be like, decision makers may use the representativeness heuristic to determine the hireability of a candidate based on how similar (s)he is to these stereotypes. If the stereotypes are not valid predictors of future success, decisions based on the representativeness heuristic will be biased.

Besides cognitive limitations, overconfidence in human judgments also contributes to the underperformance of clinical methods. Most people object to mechanical methods due to implicit belief that people gain expertise in making judgments and this expertise can improve subsequent predictions (Highhouse, 2008). Indeed, experts often have valuable insights and they can take into account idiosyncrasies of a given situation. For instance, when making judgments experts tend to rely on broken-leg cues – cues that are rare but highly diagnostic (Meehl, 1954). However, these unique cases rarely occur, are not completely reliable, and their importance is often overestimated. Empirical data has shown that overreliance on expertise may undermine the accuracy and quality of judgments (Camerer & Johnson, 1991; Dawes, Faust, & Meehl, 1989).

Efforts to Improve Decision Accuracy

Although extensive research has shown that statistical data combination methods are more accurate than human intuition, most practitioners are reluctant to rely on statistical models when making decisions. Both individual level and organizational level factors contribute to the resistance. For instance, decision makers resist statistical methods due to belief in human intuition (Highhouse, 2008) and lack of knowledge and education of statistical models (Vrieze & Grove, 2009). Moreover, Nolan and Highouse (2014) suggested that practitioners' intention to use a particular hiring practice is affected by their beliefs of whether the practice fulfills their need for autonomy. When compared with intuition-based methods, statistical data combination methods reduce the amount of autonomy that decision makers have in the decision-making process, and the decrease in autonomy reduces their intention to use statistical methods. At a broad level, politics and culture in organizations also discourage the use of purely statistical methods and encourage the use of human intuition (Johns, 1993; Muchinsky, 2004; Terpstra & Rozell, 1997).

Given the common resistance to using mechanical methods in hiring decisions (Highouse, 2008; Nolan & Highhouse, 2014), it is important to understand how to improve the accuracy of human judgment. One intervention that has received the most attention is feedback. To understand the effect of feedback, a fair amount of studies used the multiple-cue probability learning (MCPL) paradigm, where participants were asked to predict an outcome using a number of cues and learned to improve their accuracy with feedback over multiple trials (e.g., Hammond, Summers, & Deane, 1973). Extensive evidence has shown that task information feedback (information about the actual relationships between cues as well as between cues and the criterion) increases accuracy, but outcome feedback (providing the actual criterion value after human judges attempted to predict that value) is generally not effective and may even impair performance (Balzer, Sulsky, Hammer, & Sumner, 1992; Karelaia & Hogarth, 2008).

In addition to feedback, providing judges with certain training may be helpful (Harding, 2004). For instance, Garb (1989) suggests that training on biases that could influence judgment improves accuracy. In addition, one can provide judges with information on cue weighting strategies. In an unpublished study on hiring decision-making, it was found that training on the validities of different cues impacted human judgments such that participants placed greater weight on cues with stronger validities as taught in the training (Shu, Kuncel, & Yu, 2017). Moreover, they showed that providing participants with the unit-weighted composites of cues and informing them that unit-weighted composites generally approximate the overall prediction also led participants to apply more uniform weights on different cues. This latter finding indicates that providing judges with statistically combined predictions (i.e., the clinical synthesis) can influence their weighting policy to some extent.

Clinical Synthesis

The term of clinical synthesis was first proposed by Sawyer (1966) as an extension to the "sophisticated clinical" approach (Holt, 1958), which synthesizes clinical and statistical methods. The fundamental idea of the clinical synthesis is to provide the human judge with the statistical prediction along with the rest of the data and let the human judge make the final judgment. Sawyer (1966) found that, based on a review of three studies, the clinical synthesis was superior to the clinical method but the mechanical judgment was better than either one. A more recent meta-analysis of clinical judgment studies over 56 years showed that clinical methods were less accurate than statistical methods regardless of whether clinicians had access to the statistical formulas (Ægisdóttir et al., 2006). It should be noted that clinicians were provided with statistical formulas in only 5 out of the 48 studies analyzed and the meta-analysis did not directly compare the accuracy of clinical judgments with or without access to the statistical formula. Given the scarcity of prior research, I next briefly review the specific results from the six studies (reviewed by Sawyer, 1966 or Ægisdóttir et al., 2006) that directly compared the accuracy of the clinical synthesis with the clinical or mechanical prediction.

Table 1 summarizes the essential elements of each of those six studies: criterion, sample (judges), statistical information given to judges, accuracy statistic, and accuracy of clinical method, mechanical method, and clinical synthesis. The first three studies concerned judgments in clinical settings. Leli and Filskov (1981) provided six clinical psychologists and six graduate students in a clinical psychology program with tests, protocols and demographic data of brain-impaired and non-impaired individuals and asked them to identify and describe the nature of brain impairment. They found that when given the linear discriminant functions along with classifications based on each function, judges were significantly better at identifying and determining the extent of brain impairment than when the statistical information was not available. The overall hit rates improved from 41% in the clinical judgment condition to 50% in the clinical synthesis condition. However, statistical functions were still superior to the clinical synthesis (62%).

Leli and Filskov (1984) conducted a similar study where five clinical psychologists and five pre-doctoral interns in clinical psychology were asked to identify intellectual deterioration associated with brain damage based on test protocols and demographic data of brain-impaired and non-impaired individuals. Participants first made judgments without any statistical information and then rated the same protocols given a discriminant function developed by Leli and Filskov (1979) along with classifications based on the function. It was found that access to the statistical information only improved the classification accuracy of clinicians (from 58.3% to 75%) but not interns. Discriminant function was associated with the highest classification accuracy (83.3%). This indicated that compared with interns, experienced clinicians were more open to and more effective at integrating the statistical information into their final judgment. It is possible that clinicians possessed more knowledge and experience of how to interpret and use the discriminant function than the interns.

In the study by Perez (1976), six graduate students in clinical psychology made distinctions between test protocols belonging to men convicted of first and second degree murder and men convicted of crimes against property. The finding was that providing clinical judges with formulas from a multivariate analysis improved the overall categorization rate from 51% to 58%, however, it did not reach the accuracy of the multivariate function (83%).

Contrary to the finding that the clinical synthesis improved judgment accuracy, the next three studies showed that the clinical synthesis was no better than the clinical or mechanical method. Two studies examined predictions of academic performance. Melton (1952) tested whether college counselors who knew the actuarial predictions before making their clinical predictions could make better predictions of first year honor point ratio (i.e., grade point average) for freshman students they interviewed than those who did not know the actuarial predictions. Six counselors were given an actuarial table (where one can identify the predicted criterion value by finding the intersection of predictor values along the column and the row) and instructed to estimate the actuarial predictions before making their own predictions. A control group of six counselors were instructed to make predictions without any access to the actuarial table. Melton found that counselors' judgments were not substantially improved by knowing the actuarial predictions ahead, nor were they comparable with the accuracy of the actuarial method. It is important to note that in this study the actual calculated value was not provided to the judge. Instead they were given a look up tool. Similarly, Watley and Vance (1964) found that providing 111 counselors with three types of statistical data (e.g., expectancy table, correlation matrix, and multiple regression equation) did not increase the accuracy of counselors when predicting college achievement. The authors speculated that the counselors might be provided with too much information that they might not even know how to use the statistical data. Like the previous study, judges were not given a mechanical combination but tools to obtain one based on their own effort. The last of the three studies dealt with performance forecasting. In the domain of predicting college football games, Harris (1963) found that when eight football coaches, as well as the judge who developed a mathematical formula to predict outcomes of football games, were asked to improve on the predictions of the mathematical formula using any supplementary information available; they could not exceed the accuracy of the mathematical formula, which was based solely on a single index of past performance. Note that in this study there was not a comparison between unaided and aided judges. Instead the study examined whether judges could improve on a formula in this setting.

In addition to the six studies discussed above, a few more studies were located in which human judges were provided with statistical information prior to making judgments. When making the diagnosis of neurosis or psychosis based on patients' Minnesota Multiphasic Personality Inventory (MMPI) profiles, the accuracy of human judges (10 non-psychologists who were unfamiliar with the MMPI and 10 psychology graduate students who had some familiarity with the MMPI) was stably improved when they were provided with the numerical predictions from a statistical formula (Goldberg, 1968). The average accuracy of judges was a bit below the accuracy of the formula. However, simply providing judges with the formula showed a rapid increase in accuracy but the improvement in accuracy diminished over the eight-week training period. It should be noted that in the study judges were informed about the accuracy of the formula and encouraged to improve on the formula prediction. This differs from most studies where statistical predictions were merely presented as a reference.

One may argue that it is of no practical importance to compare human judgment with an optimal weighting method (e.g., linear regressions) as optimal models are rarely available in non-laboratory settings. Yet, evidence suggests that providing judges with predictions of bootstrapping models (i.e., models based on human judgment) also helps. Peterson and Pitz (1986) found that providing 20 MBA students with the predictions from a bootstrapping model derived from their previous predictions increased their subsequent predictions of the number of games a National League baseball team won during a year (absolute error dropped from 195.05 to 174.9 after providing the bootstrapping predictions), although they were not as accurate as the model itself (absolute error = 162.65). They also showed that this increase in accuracy was partially due to increased consistency of participants in using the cues available.

Additional evidence came from research on judgmental forecasts of time series data in areas such as marketing and sales. These areas are characterized by regular established patterns that are occasionally affected by exogenous factors. One potential advantage of human judgment is that judges can take into account effects of the exogenous factors, where reliable statistical estimations are difficult to reach given the infrequency of such factors (Goodwin & Fildes, 1999). However, there are mixed findings of whether judgmental adjustments of statistical predictions can consistently outperform the statistical models (Carbone, Anderson, Corriveau, & Corson, 1983; Lim

& O'Connor, 1995). For example, in the context of sales forecasting, judgmentally revising sales forecasts produced by a quantitative model occasionally but not consistently yielded forecasts that were superior to the statistical forecasts (Mathews & Diamantopoulos, 1989).

To summarize findings of prior research, several studies showed that the clinical synthesis, which is akin to providing an anchor, led to increased accuracy when compared with a purely clinical prediction (without knowledge of the statistical predictions), but the magnitude of improvements was typically small. Even when the clinical synthesis improved accuracy, it still could not reach the accuracy of mechanical methods. However, one study showed that when providing clinicians with relevant information along with a mechanical index based on a subset of the information, they predicted equally well as the mechanical index (Goldstein, Deysach, & Kleinknecht; 1973). Since the mechanical index is only based on a subset of information, it is arguable that the finding might be primarily due to clinicians having more data than the mechanical counterpart.

However, caution should be taken in drawing conclusions given the following issues of previous studies. First, all the studies were based on small convenience samples (the sample size was below 10 for the majority), which not only limits the power to detect true differences between methods but also makes it hard to generalize findings to other settings. Second, the majority of the studies involved judgments in the domain of mental health or brain impairment. Only two studies most relevant to the current study examined predictions of academic performance. It is unknown whether the nature of hiring decision-making would prevent generalizing findings in these domains to hiring decision-making. Lastly, in prior studies, human judges were typically provided with the statistical functions although there were a few studies in which judges were given the precise statistical predictions. One could argue that directly providing a precise statistical prediction may be more compelling and easier to process than providing a statistical formula, as judges may refuse to or not know how to use the formula to obtain a statistical prediction. More research is needed before firm conclusions can be made about the effectiveness of the clinical synthesis.

As discussed earlier, practitioners' resistance towards analytical methods in employment decisions might have weakened the effectiveness of the clinical synthesis. To improve upon the clinical synthesis, this study examines whether the anchoring effect could be leveraged to improve hiring decision-making. Human judges were provided with statistically combined predictions prior to making their own predictions, but unlike the clinical synthesis, where judges are aware of the statistical nature of the predictions, information about the source of the predictions was manipulated to vary along a holistic-analytical continuum (i.e., a statistical model, expert judgment, or a combination of a statistical method and expert judgment). If the reliance on intuition-based instead of analytical approaches is so pervasive, judges will be more likely to accept predictions from a more intuition-based source, and therefore show stronger anchoring. This will in turn affect their decision policy and accuracy.

Individual Differences

Not only can features and presentations of anchors influence the magnitude of anchoring effects, but also individuals with certain characteristics may be more prone to anchoring. Next, I review existing research on the role of individual differences in susceptibility to anchoring. The discussion is focused on a group of individual differences that are examined in this study.

Expertise/Experience

Theoretically, expert judges are more knowledgeable and less uncertain about making judgments in a specific domain, and thus should be less susceptible to anchoring. However, empirical findings have been mixed. On one hand, strong anchoring effects among expert judges have been demonstrated in multiple domains, such as car-selling (Mussweiler et al., 2000), real-estate price estimation (Northcraft & Neale, 1987), and legal judgments (Englich & Mussweiler, 2001).

However, anchoring may not always occur. Wilson et al. (1996) found anchoring did not occur with people who identified themselves as knowledgeable about the target judgment (a general knowledge question about the number of countries in the United Nations). In their study, individuals who identified themselves as knowledgeable of the question might know the correct answer (i.e., they knew exactly how many countries there are in the United Nations). Instead, expert judges in the other studies might only have more general experience and knowledge about the domain but not the specific knowledge of the true value. For instance, experts in the car business may be knowledgeable of how certain characteristics of the car affect the selling price but they do not necessarily know the exact price of a given car to be evaluated. As a result, expertise only matters if one knows the true value of the target judgment.

Despite the predominant view that susceptibility to anchoring is not affected by expertise, a recent study suggests that knowledge level matters for anchoring (Smith, Windschitl, & Bruchmann, 2013). For example, in the first experiment, Smith et al. found that greater knowledge of football led to smaller anchoring effects on football-related judgments. Importantly, knowledge was measured with an objective knowledge quiz, and therefore reflected true knowledge rather than confidence in their estimates. More interestingly, in a subsequent experiment, researchers manipulated the knowledge level of participants by providing them with either irrelevant or relevant information prior to making judgments. They found that anchoring was reduced when participants were given and encouraged to process information that would increase their knowledge level for the target judgment. This was in contrast with Englich (2008) where a similar manipulation failed to reduce anchoring effects by presenting participants with relevant information before making an anchored judgment. Smith et al. argued that simply providing relevant information was not sufficient to mitigate anchoring, but rather participants should be encouraged to actively process the information.

To reconcile the inconsistencies, one possibility is that expertise is more likely to reduce anchoring effects for judgments involving factual knowledge where objectively correct answers exist (e.g., football-related questions in Smith et al., 2013), compared to judgments with more subjectivity (e.g., sentencing decisions in Englich & Mussweiler, 2001). Moreover, expertise matters more to the extent that more specific knowledge is accessible. For instance, Smith et al. (2013) provided participants with relevant information to increase their knowledge level immediately before making judgments. In this situation, specific knowledge (prices of five different cars) was relatively accessible and influenced the subsequent judgment (average price of a new midsized sedan), and therefore reduced the effect of a random anchor. However, with other studies described earlier, one could argue that experts might only possess general knowledge of the domain, which could not be easily applied to the specific judgment.

To conclude, the majority of research suggests that expertise does not substantially reduce anchoring effects. A few studies showed that the impact of expertise was more profound when the target judgment concerned general knowledge with objectively correct answers (Smith et al., 2013; Wilson et al., 1996). Although experts are not immune to anchoring, one study suggests expertise may moderate the effect of mood on anchoring. Englich and Soder (2009) showed that emotions only affected the magnitude of anchoring for non-experts: experts adjusted their judgment toward the given anchor regardless of their mood, while non-experts were only susceptible to the anchoring effect when they were in a sad rather than a happy mood. Some researchers have speculated that when experts are presented with an anchor, they engage in more thorough information processing, and therefore more anchor-consistent information is activated (Furnham & Boo, 2011). Although experts have more overall knowledge (both anchor-consistent and anchor-inconsistent) about the target, increased accessibility of anchor-consistent information overrides anchor-inconsistent information, which leads to biased judgments (Smith et al., 2013). To make matters worse, experts are typically more confident and certain about what they know and may consider themselves as less susceptible to anchoring, which could cause insufficient adjustment from the anchor value (Englich & Mussweiler, 2001).

Cognitive Ability

Mixed results were found regarding the relationship between cognitive ability and anchoring susceptibility. Stanovich and West (2008) showed that cognitive ability was unrelated to some cognitive biases, including anchoring effects. In partial support of this, Oechssler, Roider, and Schmitz (2009) reported that cognitive ability predicted conjunction fallacy and conservatism in updating probability, but not anchoring. Although the effect on anchoring was not statistically significant, those with higher cognitive ability scores showed higher susceptibility to anchoring. They speculated that individuals with higher cognitive ability scores might be more likely to understand the intention of the question, and therefore provided an estimate closer to the anchor.

Other studies have demonstrated the opposite effect of cognitive ability on anchoring. Bergman, Ellingsen, Johannesson, and Svensson (2010) found that anchoring effects were reduced in a group of participants with higher cognitive ability, but the anchoring effects were still sizable. In their study, participants were asked to write down the last two digits of their social security number (anchor), and then estimated the maximum price they were willing to pay for a product. Bergman et al. used two measures of cognitive ability: the Cognitive Reflection Test (CRT; Frederick, 2005), and a more standard general intelligence test (Sjöberg, Sjöberg, & Forssén, 2006). To reconcile inconsistencies with prior findings, Bergman et al. discussed two methodological issues of the two studies mentioned above. First, Stanovich and West (2008) used self-reported SAT scores as the measure of cognitive ability, which has been shown to have non-negligible measurement errors (Kuncel, Crede, & Thomas, 2005). Second, Oechssler et al. (2009) used the CRT to measure cognitive ability. The CRT is designed to distinguish between more impulsive and more reflective decision makers, and thus measures limited aspects of cognitive abilities. In line with this, Bergman et al. found that differences in anchoring susceptibility between low and high cognitive ability groups were only observed when using the general intelligence test but not the CRT. However, a recent study failed to find any impact of cognitive ability on anchoring effects even though cognitive ability was measured with a general intelligence test and a verbal reasoning test (Furnham et al., 2012).

Although most research has shown that cognitive ability does not significantly reduce anchoring effects, some suggests that it may help individuals become less susceptible to anchoring over time. In a study where participants estimated how likely they were to win in a card-game over 140 trials, it was found that cognitive ability was unrelated to anchoring at the first quarter of trials. However, individuals with higher cognitive ability were more likely to become less susceptible to anchoring as the task progressed (Welsh, Delfabbro, Burns, & Begg, 2014). The authors inferred that instead of directly reducing anchoring effects, cognitive ability affects the rate at which individuals gain expertise and knowledge of the task, and the increase in expertise over trials explains the differences in anchor susceptibility between low and high cognitive ability groups over time.

Decision Styles

Decision styles measure the extent to which individuals use a certain strategy or style when making decisions, and therefore may be related to whether individuals are likely to be influenced by anchors during decision-making. For instance, a rational style is characterized by deliberate thoughts and logical evaluations of alternatives when making decisions (Scott & Bruce, 1995). Because of an emphasis on logic and rationality, individuals high on the rational style may be more likely to thoroughly evaluate the entire decision scenario instead of focusing on anchors, and may be less susceptible to anchoring. Only two studies directly examined the association between decision styles and anchoring effects. Similar to their findings of cognitive ability, Welsh et al. (2014) showed that decision styles did not predict susceptibility to anchoring at the beginning of the task, but was positively related to the extent to which individuals became less susceptible to anchoring over time. More specifically, participants high on preference for rationality and aptitude for rationality were more likely to become less susceptible to anchoring over trials. In other words, individuals who prefer to be rational or have the ability to be rational while making decisions tend to show reduced susceptibility to anchoring over time.

On the other hand, Cheek and Norem (2017) found that individuals with a more holistic thinking style (i.e., more attention to context and the field) were less susceptible to anchoring effects than those with a more analytic thinking style (i.e., more attention to specific elements). They speculate that individuals with an analytic thinking style might focus more on individual elements of the task, and therefore might process anchors in more depth than those with a holistic style, and more thorough processing of anchors increases anchoring effects (Bodenhausen, Gabriel, & Lineberger, 2000).

Findings of the two studies seem to be contradictory given that the rational style is typically considered as interchangeable with the analytic style (Epstein, Pacini, Denes-Raj, & Heier, 1996). The rational style is characterized by deliberate thoughts and logical evaluations of alternatives (Scott & Bruce, 1995). Similarly, the analytic style is characterized by attention to individual elements and integrating elements with detailed calculations and decision rules (Sjöberg, 2003). Therefore, further research is needed to understand the role of rationality in susceptibility to anchoring.

Resistance to Scientific Personnel Selection

Although not directly related to anchoring effects, the common resistance to using statistical methods in employment decisions raises the question, "Are individuals more resistant to statistical methods less susceptible to anchors, especially when anchors are from statistical models?". Nolan, Carter, and Dalal (2016) found that decision makers were perceived as less responsible for the outcomes of employment decisions when the statistical rather than holistic data combination method was used. Furthermore, this belief, that stakeholders would perceive decision makers as having less control and responsibility over the hiring process, affected decision makers' intention to use more standardized employee selection methods (e.g., statistical data combination method) due to perceived threat of technological unemployment. This finding suggests that the effectiveness of the clinical synthesis may depend on the extent to which decision makers are resistant to scientific personnel selection. Those who are more resistant are less likely to be influenced even when statistical predictions are presented to them.

The present study examines how those individual difference variables (i.e., prior experience with hiring decision-making, cognitive ability, decision styles, and resistance to scientific personnel selection) are associated with anchoring susceptibility. Previous literature has shown mixed findings of whether anchoring effects are reduced for experts. It is plausible that individuals with more prior experience in employee selection are more likely to make judgments based on their own strategies formed in previous hiring decisions instead of relying on the anchors provided.

Although prior research suggests that cognitive ability does not substantially reduce anchoring, it has been argued that specific abilities, instead of general cognitive ability, may be more predictive of anchoring susceptibility. This study examines whether the specific ability that is more relevant to the judgment (i.e., numerical reasoning) and the specific ability that is less relevant to the judgment (i.e., verbal reasoning) are differentially related to anchoring susceptibility. If a more relevant specific ability is more predictive of anchoring susceptibility, it may explain why past studies using general measures of cognitive ability failed to find any effects of cognitive ability on anchoring.

Little research has been focused on the relationships between decision styles and anchoring susceptibility. A few studies have shown that certain decision styles, such as the rational style, are related to decision-making outcomes such as general decision-making competence (Bruine de Bruin, Parker, & Fischhoff, 2007) and decision quality (Hough & Ogilvie, 2005). This study examines how decision styles are associated with anchoring susceptibility using the General Decision-Making Style (GDMS, Scott & Bruce, 1985) that evaluates five styles: rational (tendency to use rationality when making decisions), avoidant (tendency to avoid or delay decision-making), dependent (tendency to rely on directions from others when making decisions), intuitive (tendency to use feelings and emotions when making decisions), and spontaneous (tendency to get through the decision-making process as soon as possible).

The last individual difference variable, resistance to scientific personnel selection, is not directly related to anchoring susceptibility, but may play an important role in the effectiveness of eliciting a favorable anchoring effect. Individuals who are more resistant to scientific personnel selection are less likely to use more standardized employee selection methods (e.g., statistical data combination method), and therefore the clinical synthesis may be less effective among those individuals. However, if the resistance is only towards statistical methods, then participants who are more resistant will be more likely to accept the anchors, and therefore show stronger anchoring when being informed that the anchor source is expert judgment instead of a statistical model.

Research Questions and Hypotheses

The primary purpose of this study is to examine the effectiveness of eliciting a favorable anchoring effect to improve the accuracy of human judgment. More specifically, four research questions are examined. The first question concerns whether providing human judges with predictions from an external source, which serve as anchors upon which one can adjust their judgments, influences the decision policy and the accuracy of human judgment in the context of hiring decision-making. Based on prior research, I hypothesize that participants receiving anchors prior to making judgments will show higher accuracy than those not receiving anchors (*Hypothesis 1*).

The second question focuses on whether features of the anchor and its presentation affect the magnitude of anchoring and the accuracy of human judgment. The first feature examined is anchor source. In the current study, participants were randomly assigned to receive the information that the external prediction (i.e., the anchor) is from 1) a statistical model (i.e., the mechanical condition), 2) expert judgment (i.e., the clinical condition), or 3) a combination of a statistical method and expert judgment (i.e., the hybrid condition). The three sources of anchors fall along a holistic-analytical continuum with the hybrid in between. Given the common objection to using purely analytical methods in hiring decisions, I hypothesize that participants in the clinical condition will show the strongest anchoring effect (i.e., their judgments will be more aligned with the provided anchors), followed by those in the hybrid condition, and those in the mechanical condition will show the least anchoring (*Hypothesis 2a*). Because the actual anchor values were determined with a unit-weighting approach and unit-weighted composites of predictors typically yield larger validities than subjective judgments, I hypothesize that stronger anchoring will result in higher judgment accuracy in the clinical condition, followed by the hybrid and the mechanical conditions (*Hypothesis 2b*).

The second feature of anchors examined is anchor precision. Prior research has shown that anchor precision influences the magnitude of anchoring such that, compared with a rounded anchor, a precise anchor results in less adjustment, and thus a stronger anchoring effect (Janiszewski & Uy, 2008). In this study, participants receiving anchors were randomly assigned to receive either the high-precision anchor, a judgment value with one decimal point (e.g., 92.2), or the low-precision anchor, a judgment value to the nearest 10th with no decimal point (e.g., 90). I hypothesize that high-precision anchors will induce stronger anchoring effects (*Hypothesis 2c*) and thus higher judgment accuracy (*Hypothesis 2d*) than low-precision anchors.

The third question focuses on how individual differences predict susceptibility to anchoring. More specifically, the following individual difference measures are examined: 1) prior experience with employee selection, 2) cognitive ability, 3) decision styles (i.e., rational, avoidant, dependent, intuitive, spontaneous styles), and 4) resistance to scientific personnel selection methods. Based on prior literature, I hypothesize that 1) individuals with more prior experience in employee selection will show weaker anchoring as they are more likely to rely on their own knowledge and strategies formed in past decisions (*Hypothesis 3a*); 2) individuals higher on the rational style will show weaker anchoring as they tend to engage in logical evaluations and deliberate thoughts when making decisions (*Hypothesis 3b*); 3) individuals higher on the dependent style will show stronger anchoring given their tendency to search for advice and direction from others (*Hypothesis 3c*); and 4) individuals higher on the avoidant, intuitive, or spontaneous style will show stronger anchoring, as they are all characterized by a preference for relying on hunches and feelings to get through the decision-making process quickly (*Hypothesis 3d*). No direct hypotheses are formed regarding how cognitive ability and resistance to standardized employee selection methods relate to susceptibility to anchoring due to mixed or lack of results from prior literature.

The fourth question concerns whether individual differences interact with features of the anchor when predicting the magnitude of anchoring. I hypothesize that 1) the negative association between prior experience and anchoring will be reduced in the clinical condition when compared with the mechanical condition, because experienced decision makers are more likely to show a preference for intuitive over analytical approaches, and thus are more likely to accept anchors from a more intuitive source (*Hypothesis 4a*); 2) the negative association between the rational style and anchoring will be reduced in the mechanical condition when compared with the clinical condition, because individuals high on the rational style are characterized by the tendency to use logical evaluations rather than intuitions to make decisions, and thus are more likely to accept anchors from a more analytical source (*Hypothesis 4b*); 3) the positive association between anchoring and avoidant, intuitive, or spontaneous style will be increased in the clinical condition when compared with the mechanical condition, given that they prefer using intuitions to quickly get through the decision-making process, and thus are more likely to accept anchors from a more intuitive source (*Hypothesis 4c*); and 4) the association between resistance to standardized employee selection methods and anchoring will be negative in the mechanical condition, but positive in the clinical condition given their preference for intuitive over analytical approaches (*Hypothesis 4d*).

Method

Sample

1400 participants were recruited through Amazon Mechanical Turk. The data that were provided to participants for decision-making were real employment data obtained from Sackett, Berry, Wiemann, and Laczo (2006). The data set contains self-reported Big Five personality scores (i.e., Extraversion, Agreeableness, Openness, Emotional Stability, and Conscientiousness). These scores are linked in the dataset to a measure of each worker's counterproductive work behaviors (CWB) at an organization. The measure of CWB was not provided to participants but used as the criterion to estimate the accuracy of predictions based on the five personality scores.

Study Design

This study was a 3 (Anchor Source) X 2 (Anchor Precision) between-subjects design with an additional control condition that received no anchors and was asked to use the personality scores to make a prediction. Participants were randomly assigned to one of the following conditions: 1) control condition (no anchors were provided); 2) mechanical, high-precision anchors; 3) mechanical, low-precision anchors; 4) hybrid, high-precision anchors; 5) hybrid, low-precision anchors; 6) clinical, high-precision anchors; 7) clinical, low-precision anchors. All participants were given information about the relative predictive power of each of the Big Five scales. Table 2 summarizes the main features and examples of each condition.

Measures

Prior Experience with Employee Selection. Participants completed nine items measuring their prior hiring experience (Nolan et al., 2016; see Appendix A). Participants rated their experience with various hiring practices on a 5-point scale (1 = “never”, 2 = “a few times”, 3 = “somewhat often”, 4 = “often”, 5 = “very often”).

Cognitive Ability. Participants completed a 10-item test from the International Cognitive Ability Resource Project (ICAR; Condon & Revelle, 2014). The test consists of two types of items: verbal reasoning and numerical reasoning (See Appendix B).

General Decision-Making Style (GDMS). The GDMS assesses how individuals approach decision situations, and distinguishes between five decision styles: rational, avoidant, dependent, intuitive, and spontaneous (Scott & Bruce, 1995; see Appendix C for descriptions of and items for each decision-making style). Participants indicated the extent to which they agreed or disagreed with each item using a 5-point scale (1 = “strongly disagree”, 2 = “disagree”, 3 = “neither agree nor disagree”, 4 = “agree”, 5 = “strongly agree”).

Resistance to Scientific Personnel Selection Methods. Participants completed a 20-item questionnaire developed by Weinhardt and Beck (2017) to measure their perceptions of the fairness and usefulness of algorithms and standardized tests (See Appendix D).

Demographics. Participants self-reported on the following demographic variables: gender, age, ethnic background, education level, and employment status.

Subjective Perceptions. After making predictions, participants rated how important they considered the anchors were on a 5-point scale (1 = “not important”, 2 =

"slightly important", 3 = "moderately important", 4 = "important", 5 = "very important") and whether they have based their judgments on the anchors provided on a 5-point scale (1 = "disagree", 2 = "slightly disagree", 3 = "neither agree nor disagree", 4 = "slightly agree", 5 = "agree").

Procedure

Upon providing consent to participate in the study, participants read the background information of the decision scenario they would encounter and descriptions of the data they would use to make subsequent decisions. Then, participants answered eight validity check questions[2] to determine if they have read the background information carefully (See Appendix E). Participants then predicted the extent to which each employee would exhibit counterproductive work behaviors on the job (40 predictions in total) based on the Big Five personality scores (See Appendix F for an example).

For participants who were not in the control condition, prior to making each prediction, they were randomly assigned to (depending on which condition they belonged to) receive either the high-precision anchor, a judgment value with one decimal point (e.g., 92.2), or the low-precision anchor, a judgment value to the nearest 10th with no decimal point (e.g., 90). They also received one of the following instructions depending on the condition they were assigned to: 1) the anchors were obtained by applying a pre-determined formula to combine the five personality scores (i.e., the mechanical condition), 2) the anchors were expert predictions provided by a professional assessor who had access to the five personality scores (i.e., the clinical condition), or 3) the

[2] The eighth validity check question was removed when scoring validity check scores because nearly half of the participants failed the question.

anchors were obtained by statistically modeling judgments from an expert who had access to the five personality scores and then applying the expert's model to combine the five personality scores (i.e., the hybrid condition). Despite the different instructions about the source of anchors, participants in all three anchor source conditions received the same anchor value for each employee they predicted, and the anchor value was obtained by applying unit weighting to the five predictors. Consistent with previous research the anchoring effect was induced by asking each participant to evaluate whether the provided anchor was too high, too low, or appropriate given the predictor data (i.e., the comparative judgment), and then provided their own judgment (i.e., the absolute judgment) of each employee using a scale from 0-100 (where 0 = very likely to exhibit counterproductive work behaviors and 100 = very unlikely to exhibit counterproductive work behaviors). Participants in the control condition did not receive any anchors prior to making predictions.

After making judgments of the employees, participants provided subjective perceptions of the anchors and four commonly used decision-making methods. Then, participants completed demographic questions and measures of individual differences.

Analyses

Dependent Variables

Two parameters were estimated to determine the magnitude of anchoring[3]. The absolute deviation of each judgement from the anchor was calculated, and the absolute deviation scores across 40 predictions were averaged to obtain an average deviation score

---

[3] For participants who did not receive anchors, magnitude of anchoring was calculated between participant judgments and what the anchors would have been.

for each participant. The correlation between participant judgments and anchors was also estimated for each participant. Both the average deviation score and the anchor-judgment correlation quantify the magnitude of anchoring. A greater average deviation score indicates more adjustment from the anchor and therefore a weaker anchoring effect. A higher anchor-judgment correlation indicates a stronger anchoring effect. Participants also reported their subjective perceptions of the anchors presented to them (i.e., importance and use of anchors).

To model the decision policy of each participant, the framework of Brunswik's (1952) Lens Model was used. According to the Lens Model, human judgments and values of the environmental criterion can be described as linear combinations of environmental cues. By comparing these two linear models, one can evaluate the nature of how the human judge combines cues to yield a judgment and whether the human judgment policy corresponds to the model of the environment. Moreover, human judgment is predictive of the environmental criterion to the extent that the human judge utilizes cues in a similar way as the environment. Therefore, judgmental accuracy (r_a), can be decomposed by the Lens Model equation (Tucker, 1964, p. 528):

$$r_a = GR_eR_s + C\sqrt{(1 - R_e^2)(1 - R_s^2)}$$

In the equation, judgmental accuracy (r_a) can be determined by correlating the human judgment with the environmental criterion. Linear knowledge (G) measures the extent to which predictions of the model of the judge match predictions of the model of the environment, and can be computed by correlating predicted judgments from the model of the judge with predicted criterion values from the model of the environment. Response consistency (R_s) measures how consistently the human judge uses cues, and can be computed by correlating predicted judgments from the model of the judge with actual judgments. Environmental predictability (R_e) measures how well the criterion could be predicted with a linear model of the environmental cues, and can be computed by correlating predicted criterion values from the model of the environment with the actual criterion values. The last parameter, residual correlation (C)[4], measures the correlation between the residuals of the models of the environment and the judge, and thus captures the shared part between the human judgment policy and the environmental structure that could not be captured by the linear models.

Because anchors were estimated with the unit-weighting approach, it is possible for participants to outperform the anchors when predicting the criterion (i.e., CWB). To assess the performance of each participant relative to the anchors, the accuracy of anchors was first determined by correlating anchors with the criterion values, and then the relative accuracy of each participant was estimated by subtracting the accuracy of anchors from the accuracy of participant judgments. A positive value indicates the participant outperforms the anchors and a negative value indicates the participant underperforms the anchors. Further analyses used judgment accuracy, relative accuracy, linear knowledge, and response consistency as dependent variables. Hypotheses were only formed with judgment accuracy and analyses with relative accuracy, linear knowledge and response consistency were exploratory.

To capture the weighting policy of each participant, participant judgments were regressed on the five personality scores at the participant level and standardized

[4] Further analyses were not conducted on residual correlation due to negligible magnitudes (the average of residual correlation in the whole sample is .03).

regression coefficients of each personality score were obtained. To ensure weights were comparable across participants, the relative weight of each personality score was estimated by dividing the standardized regression coefficient by the sum of the coefficients across all five predictors. The standard deviation of predictor weights was also computed to assess weight variation.

Analyses

To test whether providing human judges with predictions from an external source would affect the decision policy and the accuracy of their judgments, independent-groups t-tests were conducted comparing estimates of magnitude of anchoring, each Lens Model parameter (i.e., judgment accuracy, linear knowledge, response consistency, environmental predictability[5]), and the weighting policy between participants who received anchors with those in the control condition. Multiple regressions were performed to test the effects of anchor features (i.e., source and precision) on the magnitude of anchoring and Lens Model parameters.

Pearson product-moment correlation coefficients were computed to assess the correlation of each individual difference variable with the magnitude of anchoring and Lens Model parameters. Moderated regressions were conducted to test the interactions of individual differences with anchor source. Each dependent variable (i.e., average deviation, anchor-judgment correlation, and judgment accuracy) was regressed on anchor source, the individual difference variable (mean centered), and the interaction term.

[5] Differences in environmental predictability were tested to ensure comparable task predictability between conditions.

Results

Data Cleaning

The initial data set was cleaned based on the following criteria. First, 12 participants were removed because they did not complete the study or provided text instead of numerical responses for the judgment task and thus were not interpretable. Then the comparative judgment was compared with the absolute judgment to assess level of correspondence. When the option “too high” was selected for the comparative judgment, and the deviation of the absolute judgment from the anchor was negative (i.e., they adjusted downward as they thought the anchor was too high given the predictor values), correspondence was coded as 1 and other responses were coded as 0. When the option “too low” was selected, and the deviation of the absolute judgment from the anchor was positive (i.e., they adjusted upward as they thought the anchor was too low given the predictor values), correspondence was coded as 1 and other responses were coded as 0. When the option “just appropriate” was selected, and the absolute deviation of the absolute judgment from the anchor was less than 5 points, correspondence was coded as 1 and other responses were coded as 0. Then correspondence scores were averaged across the 40 predictions made to determine the level of correspondence for each participant and a low level of correspondence indicates careless responding. 144 participants who had a low level of correspondence (< .81)[6] were excluded from the data set. Additionally, 10 participants completed the study in less than 14 minutes[7] and did not

---

[6] This cutoff was determined by Q1-1.5IQR, where Q1 = .925 was the first quartile of correspondence and IQR = .075 was the interquartile range of correspondence. Scores lower than this cutoff were considered as outliers.
[7] This cutoff was 1/3 of the median of the completion time, which was considered as the lower-bound of the legitimate time required to complete the study.

get full credit for validity check questions, and therefore were excluded. The final data set consisted of 1,234 participants (52.9% females, $M_{age} = 39.24$) who provided valid responses.

Effects of External Predictions

Results of independent groups t-tests indicated that participants who received anchors (*M* = 10.14, *SD* = 3.93) showed significantly lower average deviation score than the control condition (*M* = 17.49, *SD* = 5.66), $t(222.05) = -17.14$, $p < .001$. Differences in anchor-judgment correlation were also significant between participants who received anchors (*M* = .87, *SD* = .11) and the control condition (*M* = .68, *SD* = .27), $t(198.92) = 9.57$, $p < .001$. Therefore, anchoring was successfully induced in the present study.

Participants who received anchors (*M* = .41, *SD* = .14) showed significantly higher accuracy than the control condition (*M* = .38, *SD* = .20), $t(220.76) = 2.28$, $p = .02$. Therefore, hypothesis 1 was supported. Participants who received anchors (*M* = .94, *SD* = .07) also showed significantly higher response consistency than the control condition (*M* = .89, *SD* = .13), $t(205.64) = 5.26$, $p < .001$. Differences in linear knowledge between participants who received anchors (*M* = .76, *SD* = .18) and the control condition (*M* = .71, *SD* = .33), $t(209.15) = 1.89$, $p =.06$ were in the hypothesized direction. Differences in environmental predictability were not significant between participants who received anchors (*M* = .45, *SD* = .16) and the control condition (*M* = .45, *SD* = .15), $t(1232) = -.37$, $p = .71$ as was expected due to the randomization of stimulus material across conditions. Figure 1 illustrates how judgment accuracy, response consistency, and magnitude of anchoring differ between participants who received anchors and those who did not receive anchors.

Participants who received anchors (*M* = .12, *SD* = .10) placed greater weight on Extraversion than the control condition (*M* = .00, *SD* = .61), *t*(189.96) = 2.81, *p* = .01. Participants who received anchors (*M* = .12, *SD* = .11) also placed greater weight on Openness than the control condition (*M* = .03, *SD* = .41), *t*(192.76) = 2.84, *p* < .01. Participants who received anchors (*M* = .30, *SD* = .14) placed less weight on Conscientiousness than the control condition (*M* = .46, *SD* = .33), *t*(200.47) = -6.24, *p* < .001. When comparing weight variation, those who received anchors (*M* = .12, *SD* = .14) showed less variation of weights across predictors than the control condition (*M* = .28, *SD* = .43), *t*(195.49) = -5.07, *p* < .001. There were not significant differences in the weight placed on Agreeableness between participants who received anchors (*M* = .23, *SD* = .22) and the control condition (*M* = .23, *SD* = .34), *t*(217.69) = -.12, *p* =.91. Differences of weight on Emotional Stability were also not significant between participants who received anchors (*M* = .23, *SD* = .14) and the control condition (*M* = .28, *SD* = .39), *t*(197.19) = -1.93, *p* =.06.

Effects of Anchor Features

Table 3 shows results of regressing the magnitude of anchoring on anchor source and anchor precision. Precision significantly predicted the average deviation score, β = -.07, *t*(1041) = -2.41, *p* = .02, indicating that participants showed stronger anchoring when anchors were presented at a precise level. Neither source nor precision had significant effects on anchor-judgment correlation. Therefore, hypothesis 2a was not supported. Hypothesis 2c was partially supported when anchoring was measured as average deviation score. Source significantly predicted importance and use of anchors. Specifically, participants in the clinical condition (β = -.13, *t*(1041) = -3.63, *p* < .001) and those in the hybrid condition (β = -.08, t(1041) = -2.18, p = .03) perceived anchors as less important than those in the mechanical condition. Participants in the clinical condition (β = -.09, t(1041) = -2.61, p = .01) also reported less use of anchors than those in the mechanical condition.

Table 4 shows results of regressing each of the Lens Model parameters on source and precision. The interaction of clinical x precise was significant when predicting judgment accuracy, β = -.12, t(1039) = -2.14, p = .03, relative accuracy, β = -.14, t(1039) = -2.52, p = .01, linear knowledge, β = -.13, t(1039) = -2.35, p = .02, and response consistency, β = -.11, t(1039) = -2.03, p = .04. The clinical x precise interaction for accuracy is plotted in Figure 2. The clinical x precise interaction for relative accuracy is plotted in Figure 3. The clinical x precise interaction for linear knowledge is plotted in Figure 4. The clinical x precise interaction for response consistency is plotted in Figure 5. Simple slope analyses were conducted to interpret the interactions. Results are shown in Table 5. When anchors were precise, participants who received clinical anchors showed lower accuracy, β = -.13, t(514) = -2.52, p = .01, lower relative accuracy, β = -.13, t(514) = -2.55, p = .01, and lower linear knowledge, β = -.13, t(514) = -2.48, p = .01, than those who received mechanical anchors. When anchors were rounded, participants who received clinical anchors showed higher consistency than those who received mechanical anchors, β = .10, t(525) = 2.02, p = .04. Therefore, neither hypothesis 2b nor hypothesis 2d was supported. A summary of descriptive statistics of the magnitude of anchoring and the Lens Model parameters by condition is shown in Table 6.

Correlations between the magnitude of anchoring and Lens Model parameters are shown in Table 7. Both average deviation (r = -.43) and anchor-judgment correlation (r

= .61) were correlated with response consistency. More specifically, participants who were more consistent tended to show less deviation from anchors and their predictions were more related to anchors. Contrary to the hypothesis, only anchor-judgment correlation was weakly associated with judgment accuracy (r = .11). Exploratory analyses were conducted to further examine the relationship between the magnitude of anchoring and accuracy. Polynomial regression was performed to estimate the relationship between average deviation and relative accuracy. As shown in Table 8, adding the quadratic term significantly increased R^2 (from .00 to .03). As shown in Figure 6, when relative accuracy was negative (i.e., participants underperformed anchors), more deviation was associated with lower relative accuracy; when relative accuracy was positive (i.e., participants outperformed anchors), more deviation was associated with higher relative accuracy. Participants were then separated into two groups based on relative accuracy: outperformers (whose relative accuracy was positive) and underperformers (whose relative accuracy was negative). Among outperformers, average deviation was positively correlated with relative accuracy, $r(667) = .38$, $p < .001$ and judgment accuracy, $r(667) = .10$, $p = .01$. Among underperformers, average deviation was negatively correlated with relative accuracy, $r(374) = -.43$, $p < .001$ and judgment accuracy, $r(374) = -.17$, $p < .001$.

Effects of Individual Differences

Table 9 shows the correlations of individual difference variables with the magnitude of anchoring and the Lens Model parameters. Table 10 shows the correlations among individual difference variables. One individual difference variable was significantly correlated with the magnitude of anchoring. Specifically, participants who had more prior experience with hiring tended to score higher on average deviation score (*r* =.12), and lower on anchor-judgment correlation (*r* = -.09). Therefore, hypothesis 3a was supported. Hypotheses 3b-3d were not supported. As for subjective perceptions of anchors, participants who perceived anchors as more important tended to have more prior experience (*r* = .10), score lower on cognitive ability (*r* = -.22), and score higher on dependent (*r* =.11), intuitive (*r* = .20), and spontaneous (*r* = .07). Participants who self-reported more use of anchors tended to score lower on cognitive ability (*r* = -.22), and higher on dependent (*r* = .13), intuitive (*r* =.18), and spontaneous (*r* = .08), and perceive scientific personnel selection method as less unfair (*r* = -.07).

Two variables were significantly correlated with judgment accuracy: participants with higher judgment accuracy tended to have less prior experience (*r* = -.08), and score higher on cognitive ability (*r* = .14). Participants who showed higher relative accuracy tended to have less prior experience in hiring (*r* = -.11), score higher on cognitive ability (*r* = .20) and rational (*r* = .08), and score lower on intuitive (*r* = -.09) and spontaneous (*r* = -.09). For participants who scored higher on linear knowledge, they tended to have less prior experience (*r* = -.08), score higher on cognitive ability (*r* = .15), and score lower on intuitive (*r* = -.07) and spontaneous (*r* = -.06). All the individual difference variables except dependent style were significantly associated with response consistency: participants who were more consistent tended to have less prior experience (*r* = -.14), score higher on cognitive ability (*r* = .30), rational (*r* = .14), score lower on avoidant (*r* = -.07), intuitive (*r* = -.10), spontaneous (*r* = -.14), and perceive scientific personnel selection methods as more useful (*r* = -.07).

Table 11 shows regression results for significant interactions when predicting anchoring. Table 12 shows regression results for significant interactions when predicting accuracy. Results of hypothesized interactions were reported first, followed by exploratory analyses of other interactions. None of the hypothesized interactions of individual difference variables with anchor source when predicting anchoring were significant (i.e., Hypotheses 4a-4d). However, the interaction of cognitive ability x hybrid was significant when predicting average deviation, $\beta = -.10$, $t(1039) = -2.19$, $p = .03$. Simple slope analyses showed that cognitive ability predicted average deviation only in the mechanical condition, $\beta = .13$, $t(342) = 2.43$, $p = .02$.

When predicting accuracy, several individual differences x source interactions were significant. The interaction of clinical x experience was significant, $\beta = .10$, $t(1039) = 2.33$, $p = .02$. Simple slope analyses showed that experience predicted accuracy only in the mechanical condition, $\beta = -.19$, $t(342) = -3.58$, $p < .001$. The interaction of clinical x cognitive ability was significant, $\beta = .12$, $t(1039) = 2.63$, $p = .01$. Simple slope analyses showed that cognitive ability predicted accuracy only in the clinical condition, $\beta = .23$, $t(353) = 4.43$, $p < .001$. The interaction of clinical x unfairness was significant, $\beta = -.12$, $t(1039) = -2.66$, $p = .01$. Simple slope analyses showed that unfairness predicted accuracy only in the clinical condition, $\beta = -.12$, $t(353) = -2.29$, $p = .02$. The interaction of clinical x unusefulness was significant, $\beta = -.15$, $t(1039) = -3.30$, $p < .001$. Simple slope analyses showed that unusefulness predicted accuracy only in the clinical condition, $\beta = -.16$, $t(353) = -3.05$, $p < .01$.

Discussion

The purpose of this study is to examine the effectiveness of eliciting a favorable anchoring effect to improve hiring decision-making. Specifically, the study examines whether 1) providing predictions from an external source (i.e., anchors) will affect the decision policy and the accuracy of human judgment, 2) features of the anchors (i.e., source and precision) will affect the magnitude of anchoring and the accuracy of human judgment, 3) individual differences predict susceptibility to anchoring, and 4) individual differences interact with features of the anchors when predicting the magnitude of anchoring.

Overall, a number of important findings were obtained. First, providing external predictions as a decision aid increased consistency and accuracy of human judgment and influenced how individuals assigned weights to different predictors. Second, the magnitude of anchoring was increased when anchors were presented at a more precise level. Effects on judgment accuracy were more complicated: when anchors were presented at a precise level, individuals who were informed that the anchors were provided by an expert showed slightly lower accuracy than those who were informed that the anchors were obtained from a statistical formula. Third, the magnitude of anchoring decreased among individuals with more prior experience in employee selection. Finally, only cognitive ability interacted with anchor source when predicting the magnitude of anchoring, and several interactions emerged when predicting judgment accuracy. The following section discusses each of the findings in more detail.

The fact that participants who received anchors showed a fairly high correlation between their judgments and the anchors (mean r = .87) when compared to those who did not receive anchors (mean $r = .68$) indicates that the anchoring effect was successfully induced in the context of hiring decision-making. Different patterns in the weighting policy of individuals who received anchors and who did not lend further support to the robustness of the anchoring effect. In this study, all participants were explicitly told that Agreeableness, Emotional Stability, and especially Conscientiousness are more predictive of the outcome than the other two predictors (i.e., Extraversion and Openness). Without additional information (i.e., anchors), participants were supposed to place greater weight on those three predictors. On the other hand, those who received anchors would assign similar weight to each predictor and thus would show less variation across predictor weights because their judgments would be more aligned with the anchors, which were unit-weighted composites of the five predictors. This was supported by the results: individuals who received anchors placed greater weight on Extraversion and Openness (two predictors that are less predictive of CWB), less weight on Conscientiousness (the best predictor of CWB), and showed less weight variation when compared with the control condition. These results are likely due to greater emphasis being placed on the anchors in anchoring conditions. That is, the weight placed on Openness and Extraversion is likely indirect. Despite more emphasis being placed on less valid predictors, participants in the anchoring conditions still provided more valid predictions of CWB. The finding that participants who received anchors showed higher accuracy than those who did not receive anchors reveals that the anchoring effect, which is a robust decision heuristic that often leads to systematic biases, can be used to reduce bias of human judgment under certain circumstances.

These findings suggest that simple interventions such as providing predictions from an external source can influence how individuals integrate data to make decisions, and can increase the accuracy of their decisions. In a real hiring situation, if an entire reliance on statistical predictions is impossible due to resistance to using purely analytical approaches, providing decision makers with statistically combined predictions may help align human predictions with statistical predictions, and therefore increase judgment accuracy.

In fact, anchors also increased the consistency of human judgment. Statistical methods tend to be superior because the inconsistency of human judges in applying weights undermines accuracy. The improvement in consistency in the presence of anchors may contribute to the increase in accuracy although the effect is not overwhelming. Prior research has shown that consistency did not always have a large effect on accuracy (Bisantz et al., 2000). As will be discussed later, consistency and accuracy was only moderately related ($r = .28$), and they demonstrated different patterns of associations with other variables examined in this study. It is possible that the predictability of which individual would be more likely to exhibit counterproductive work behaviors given their personality scores is not high enough, so accuracy is relatively insensitive to consistency.

The results align with past research showing that compared with rounded anchors, precise anchors result in less adjustment and stronger anchoring (Janiszewski & Uy, 2008). It was found that presenting anchors at a more precise level (i.e., numbers with one decimal point vs. numbers rounded to the nearest 10th) induced stronger anchoring among participants. The impact of the second feature, anchor source, was more complex.

As an extension of the clinical synthesis method, this study examined whether the magnitude of anchoring would vary depending on the source of anchors. Clinical synthesis is a data combination method in decision-making where human judges are provided with a statistically combined prediction in addition to the predictor data, and judges can decide how to integrate the statistical prediction with the rest of the data. On the other hand, with a purely clinical method, judges integrate the predictor data to reach an overall prediction without knowledge of the statistical prediction. Although some evidence suggests the clinical synthesis was more accurate than the clinical method, the magnitude of improvement was typically small (Sawyer, 1966). One explanation is that with the clinical synthesis, judges explicitly know the external prediction is obtained from a statistical model and they may not give any weight to the statistical prediction due to resistance to analytical approaches.

In this study, information about the source of the external predictions was manipulated to vary along a holistic-analytical continuum. If the preference for holistic over analytical methods among most decision makers is prevalent, judges will be more likely to align their predictions with the anchors when they know the anchors are from a more holistic source. However, this was not supported by the results. Knowing that the anchors were obtained from an intuition-based source (i.e., expert judgment) did not increase the degree of anchoring among participants. In fact, contradictory findings were observed when examining the subjective measures of anchoring. Specifically, participants were more likely to self-report that they considered the anchors important and have based their predictions on the anchors when they were informed that the anchors are from an analytical source (i.e., a statistical model). As will be discussed later, different results were obtained when comparing the objective (e.g., average deviation score) and subjective (e.g., importance) indicators of anchoring, and it may signal the unreliability of self-reported measures of anchoring. These results suggest that when providing decision makers with external predictions as a decision aid, the predictions should be presented as more precise numbers to enhance the potency of the numbers. However, knowing the source of the predictions did not affect the degree to which judges adopt the predictions. In other words, decision makers are no more likely to accept a decision aid when it comes from another human expert rather than a statistical model. The finding that the source of anchors did not influence the magnitude of anchoring provides tentative support for the selective accessibility approach to anchoring, which posits that comparing the target judgment with the anchor value increases the accessibility of anchor-consistent information about the target, and this increase is not dependent on the relevance of the anchor.

The rationale for improving human judgment with anchors is that human predictions, when aligned with anchors, should be predictive of the outcome if the anchors have strong validity. In other words, the stronger the anchoring is and the more accurate the anchor is, the higher the accuracy of human judgment should be, and the effects of anchor features on anchoring should transfer to the effects on accuracy. Contrary to the hypothesis, the magnitude of anchoring was not highly correlated with accuracy, and different effects of anchor features on accuracy were observed. When examining differences in accuracy, a significant interaction between source and precision was observed: when anchors were presented at a precise level, participants who were informed that the anchors were provided by an expert showed lower accuracy than those who were informed that the anchors were obtained from a statistical formula.

One explanation for the weak association between anchoring and accuracy is that the anchors presented to participants were suboptimal predictions and it was possible for participants to outperform the anchors. Because participants were aware that three of the five predictors are more predictive than the other two, participants were likely to improve upon the anchors if they developed a weighting policy with greater weight on those three predictors. In other words, greater deviation from anchors resulting in weaker anchoring is not necessarily irrational and the relationship between anchoring and accuracy should not be linear. Exploratory analyses revealed a curvilinear relationship between anchoring and relative accuracy (i.e., the accuracy of human judgment relative to the accuracy of anchors): when participants outperformed the anchors and the more their predictions deviated from the anchors, the more improvement in accuracy they made beyond the anchors; when participants underperformed the anchors and the more their predictions deviated from the anchors, the more decrease in accuracy they had as compared to the anchors.

The curvilinear relationship between anchoring and relative accuracy then transferred to the relationship between anchoring and accuracy: greater deviation from anchors was correlated with higher accuracy among participants who showed improvement over the anchors, while greater deviation was correlated with less accuracy among underperformers. In other words, stronger anchoring led to higher accuracy only when individuals did not develop a better weighting schema than unit weighting. When individuals were able to apply a weighting schema that was more accurate than unit weighting, weaker anchoring implied reasonable adjustment and thus higher accuracy. One implication is that although human judgment generally underperforms a statistical model, some decision makers are able to improve upon the statistical model when additional information about how to develop and apply a more accurate weighting method exists.

Although anchoring did not have a strong linear association with accuracy, its correlation with response consistency was rather high. Because anchors were unit-weighted composites of the five predictors, a greater alignment between human predictions and the anchors implies a stronger linear association between human predictions and the predictors, and therefore higher consistency. In addition, effects of anchor features on consistency were not consistent with the effects on accuracy: when anchors were rounded, participants who were informed that the anchors were provided by an expert (i.e., the clinical condition) showed higher consistency than those who were informed that the anchors were obtained from a statistical formula (i.e., the mechanical condition). Recall that when anchors were precise, participants in the clinical condition showed lower accuracy than those in the mechanical condition. It should be noted that the patterns of the interaction effects were similar between accuracy (Figure 2) and consistency (Figure 5): when anchors were rounded, the clinical condition showed slightly higher consistency/accuracy than the mechanical condition, and the pattern was reversed when anchors were precise. Differences in the results of significance testing may be due to different effect sizes and the fact that consistency does not always have a large effect on accuracy. Moreover, the fact that individuals in the clinical condition showed the least improvement on the anchors when anchors were precise implies that when external predictions are presented as precise numbers and from an intuition-based source, participants might have shown the greatest acceptance of the anchors and were less likely to outperform or underperform the anchors.

In addition to the features of the anchors, this study examined whether individual differences predict susceptibility to anchoring and other aspects of decision-making. Prior experience with employee selection was the only individual difference variable that was associated with the magnitude of anchoring. Specifically, individuals who had more prior experience with hiring decisions showed weaker anchoring. This is inconsistent with the majority of prior studies showing that expert and novice judges are equally susceptible to anchoring (Englich & Mussweiler, 2001), but consistent with the hypothesis that experienced decision makers are more likely to rely on their own knowledge and strategies formed in past decisions instead of anchors when making decisions. The finding that cognitive ability was not correlated with anchoring is consistent with prior literature (Furnham et al., 2012; Stanovich & West, 2008). In addition, neither verbal reasoning nor numerical reasoning predicted the magnitude of anchoring and they showed similar patterns of associations with the dependent variables. This refutes the proposition that, at least these specific abilities, instead of general cognitive ability, may be more predictive of anchoring susceptibility. Little research has examined how decision styles predict anchoring susceptibility. Results of this study provide preliminary evidence that individuals with different decision styles, at least those measured by GDMS (Scott & Bruce, 1985), are equally susceptible to anchoring. The lack of associations between individual differences and anchoring demonstrates that the anchoring effect is quite robust and it affects the majority of decision makers regardless of their characteristics.

It should be noted that cognitive ability and three of the five decision styles (i.e., dependent, intuitive, and spontaneous) were correlated with the subjective *perceptions* of anchoring. Specifically, individuals lower on cognitive ability, or higher on dependent, intuitive, or spontaneous styles were more likely to self-report that they considered the anchors important and have based their predictions on the anchors. The discrepancies between objective and subjective indicators of anchoring reflect that self-reported measures of anchoring may not be reliable. Individuals may not be fully aware of their own decision policy or they may disguise their decision policy to be socially acceptable.

Although individual differences did not predict anchoring susceptibility, some of them were correlated with consistency and accuracy of human judgment. More experience in employee selection was associated with lower consistency and accuracy. This is consistent with decision-making research showing that having prior experience in the domain does not necessarily improve the accuracy of one's judgments, and may at times impede judgments (e.g., Garb, 1989; Oskamp, 1962). Instead of following instructions provided to them, experienced decision makers were more likely to rely on other decision rules, presumably those formed in past hiring decisions, which may not be optimal in this study. The finding that individuals higher on cognitive ability showed higher judgment consistency and accuracy provides further support that cognitive ability is one of the best predictors of many important life outcomes, even in the context of hiring decision-making.

Although not directly associated with accuracy, individuals higher on the rational style and those lower on the avoidant, intuitive, or spontaneous style tended to be more consistent. The positive association between rational and consistency is not surprising given that the rational style is characterized by deliberate thoughts and logical evaluations of alternatives, and has been found to be associated with positive outcomes, such as satisfaction with career choice (Crossley & Highhouse, 2005), high job performance (Russ, McNeilly, & Comer, 1996), and decision-making competence (Bruine de Bruin et al., 2007). On the other hand, individuals with the intuitive style tend to rely on hunches and feelings and those with the spontaneous style are characterized by a sense of immediacy and prefer to get through the decision-making process as soon as possible. Both intuitive and spontaneous styles are characterized by decisions being made relatively quickly, and thus are likely to be more error-prone and inconsistent (Russ et al., 1996). It should be noted that decision styles are not mutually exclusive, and one can choose different styles when facing different decision scenarios (Scott & Bruce, 1995).

Though it was not hypothesized, exploratory analyses revealed a significant interaction between cognitive ability and source when predicting anchoring such that higher cognitive ability was related to weaker anchoring only in the mechanical condition. This is counterintuitive given that individuals higher on cognitive ability tend to use logical evaluations rather than intuitions when making decisions, and should be more likely to accept anchors from a more analytical source. In other words, cognitive ability should be less predictive of anchoring in the mechanical condition. It is unclear what might have contributed to this interaction effect.

When predicting accuracy, several interactions between individual differences and anchor source were observed. Caution should be given when interpreting those interaction effects and further research is needed to test the generalizability of those findings. First, a significant interaction between experience and source was found such that more experience was associated with lower accuracy only in the mechanical condition. It is possible that experienced decision makers were more resistant to relying on external predictions that are obtained from a statistical source than an intuition-based source and instead they relied more on their own decision rules formed in past decisions that were not necessarily accurate. Second, a significant interaction between cognitive ability and source was found such that higher cognitive ability was associated with higher accuracy only in the clinical condition. Given the curvilinear relation between anchoring and accuracy and the fact that cognitive ability was not related to anchoring in the clinical condition, it is possible that individuals with relatively high cognitive ability tended to outperform the anchors while individuals with relatively low cognitive ability tended to underperform the anchors in the clinical condition. Even though the two groups showed similar levels of deviation from anchors, more deviation from anchors was rational among high cognitive ability individuals and thus increased accuracy while more deviation was irrational among low cognitive ability individuals and thus decreased accuracy. As a result, the effect of cognitive ability on accuracy was stronger in the clinical condition.

Finally, significant interactions were observed between anchor source and both measures of resistance to scientific personnel selection. Individuals who perceived scientific personnel selection as more unfair and unuseful showed lower accuracy only in the clinical condition. It is possible that individuals who were more resistant to statistical methods tended to underperform the anchors while individuals who were less resistant tended to outperform the anchors in the clinical condition. Even though they showed similar levels of deviation from anchors, more deviation from anchors was rational among less resistant individuals and thus increased accuracy while more deviation was irrational among more resistant individuals and thus decreased accuracy. As a result, the effect of resistance on accuracy was stronger in the clinical condition.

Limitations and Future Research

One limitation of this study concerns the nature of the anchors presented to participants. In order to understand whether individuals could improve upon the external predictions provided to them, the external predictions were estimated using a suboptimal weighting policy of the predictors (i.e., simple unit weighting). Suboptimal anchors were used to allow judges to improve on the anchors and permit the inclusion of an incentive to help ensure commitment to the task. However, one issue with using suboptimal predictions as anchors is that the relationship between the magnitude of anchoring and the accuracy of human judgment is not necessarily linear. In other words, weaker anchoring could result from either rational or irrational deviations from the anchors. Irrational deviations decrease accuracy while rational deviations increase accuracy. As a result, it is difficult to examine how the effects of anchor features and individual differences on the magnitude of anchoring will transfer to the effects on accuracy. Future research should consider using more optimal anchors, such as those obtained from linear regressions, to better understand how the accuracy of human judgment changes depending on the magnitude of anchoring.

A related issue is how the control condition was set up in the study. All participants, including those in the control condition, were informed that three of the five predictors are more predictive of the criterion. This information can be considered as task-related knowledge, and even though the control condition did not receive anchors as a decision aid, this task-related knowledge might have increased the accuracy of their predictions. Since the control group had information that would allow them to match or exceed the anchor, this served as a stricter test of anchoring. Therefore, the control condition would not serve as the lowest baseline of accuracy to be compared with those receiving anchors. Results showed that participants who received anchors showed higher accuracy than those who did not receive anchors. It is possible for the anchors to demonstrate greater improvement in accuracy without this task-related knowledge. Future research should create a control condition without any task-related knowledge about the validity of predictors to examine how anchors alone influence the accuracy of human judgment.

Another limitation concerns the nature of the participants in the study. The fact that participants were recruited through Amazon Mechanical Turk might have contributed to the relatively low level of experience with hiring decisions in this sample (1.92 out of 5). As the majority of participants did not have greater experience in hiring, they were more likely to show a stronger anchoring effect when provided with the external predictions. The lack of variation in experience could also dilute the effects of experience on decision policy and accuracy. Ideally, future research should include participants from populations with higher level of experience, for instance, human resources practitioners, and diverse levels of experience to better understand how experience affects the effectiveness of anchoring for improving hiring decision-making.

A final direction for future research is to examine the relationship between response consistency and judgment accuracy. It is unclear why consistency was not highly correlated with accuracy and the two features of anchors showed different impacts on response consistency and judgment accuracy. Future research should look into how response consistency contributes to judgment accuracy. If consistency is not highly correlated to accuracy, more research should examine what factors may account for individual differences in accuracy for the same judgment.

Conclusion

This study examined the effectiveness of eliciting a favorable anchoring effect to improve hiring decision-making. Results suggest that providing decision makers with predictions from an external source affects their decision policy and increases the accuracy of human judgment. The presentation of the anchors affects the degree of anchoring such that there is a greater degree of anchoring when anchors are presented as precise, rather than rounded numbers. However, knowing that the external predictions come from a human expert rather than a statistical model does not increase the reliance of decision makers on the external predictions.

An intriguing finding is that although extensive evidence suggests statistical data combination methods are generally superior to human judgment, some decision makers are able to develop and apply a more accurate weighting method than the statistical method when provided with task-related knowledge. As a result, the accuracy of human judgment outperforms that of the statistical model (although an intentionally less accurate one).

Finally, the anchoring effect is a ubiquitous phenomenon in human judgment as individuals with different characteristics are equally susceptible to anchoring. Moreover, judgment consistency and accuracy demonstrate greater associations with individual differences including experience, cognitive ability, and certain decision-making styles.

Overall the results suggest decision makers can be selected and then trained to be more consistent and accurate in their judgments.

Table 1. *Descriptions of Six Studies Comparing Clinical Synthesis with Clinical or Mechanical Method*

Citation	Criterion	Sample (Judges)	Statistical Information	Accuracy Statistic	Accuracy		
					Clinical Method	Mechanical Method	Clinical Synthesis
Leli and Filskov (1981)	Brain Impairment	6 clinical psychologists and 6 graduate students in clinical psychology	Linear discriminant functions along with raw scores, unstandardized lambda weights, and classifications based on each function	Hit rate	41%	62%	50%
Leli and Filskov (1984)	Intellectual Deterioration	5 clinical psychologists and 5 pre-doctoral interns in clinical psychology	Discriminant function along with raw scores, unstandardized lambda weights, and classifications based on the function	Hit rate	62.5% (Interns) 58.3% (Clinicians)	83.3%	66.5% (Interns) 75% (Clinicians)
Perez (1976)	Homicidality	6 graduate students in clinical psychology (2 first-year students, 2 practicum students, and 2 interns)	Formulas of a multivariate analysis	Hit rate	51%	83%	58%
Melton (1952)	Academic Performance	12 college counselors who were graduate students in psychology or educational psychology	An actuarial table where one can identify the predicted criterion value by finding the intersection of predictor values along the column and the row	Mean absolute error	0.45	0.35	0.45
Watley and Vance (1964)	Academic Performance	66 counselors and 45 naïve judges (primarily freshmen and sophomore students)	Expectancy table, correlation matrix, and multiple regression equation	Hit rate	66.6%	79%	66.9%
Harris (1963)	Outcomes of Football Games	8 football coaches and Litkenhous (who developed the mathematical formula)	Predictions of a mathematical formula based on past performance	Hit rate	-	66%	60% (Coaches) 63% (Litkenhous)

Table 2. *Descriptions of Experimental Conditions*

Number	Condition	Anchor Source Instruction Example	Anchor Precision Instruction Example
1	Control (no anchors)	-	-
2	Mechanical, high-precision	To guide your judgment, below is a prediction obtained by applying a pre-determined formula to combine the five personality scores.	The predicted integrity score based on the formula for Mark is: 64.8 (on a scale from 0 to 100)
3	Mechanical, low-precision	To guide your judgment, below is a prediction obtained by applying a pre-determined formula to combine the five personality scores.	The predicted integrity score based on the formula for Mark is: 60 (on a scale from 0 to 100)
4	Hybrid, high-precision	To guide your judgment, below is a prediction obtained by statistically modeling judgments from an expert who has access to the five personality scores and then applying the expert's model to combine the five personality scores.	The predicted integrity score based on the expert's statistical model for Mark is: 64.8 (on a scale from 0 to 100)
5	Hybrid, low-precision	To guide your judgment, below is a prediction obtained by statistically modeling judgments from an expert who has access to the five personality scores and then applying the expert's model to combine the five personality scores.	The predicted integrity score based on the expert's statistical model for Mark is: 60 (on a scale from 0 to 100)
6	Clinical, high-precision	To guide your judgment, below is an expert's prediction provided by a professional assessor who has access to the five personality scores.	The predicted integrity score based on the expert judgment for Mark is: 64.8 (on a scale from 0 to 100)
7	Clinical, low-precision	To guide your judgment, below is an expert's prediction provided by a professional assessor who has access to the five personality scores.	The predicted integrity score based on the expert judgment for Mark is: 60 (on a scale from 0 to 100)

Table 3. *Regressions of Magnitude of Anchoring on Anchor Features (N = 1045)*

	Average Deviation			Ancho-Judgment Correlation			Importance			Use		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Clinical	-.11	.30	-.01	.00	.01	.00	-.28	.08	-.13***	-.24	.09	-.09**
Hybrid	.10	.30	.01	.00	.01	.00	-.17	.08	-.08*	-.15	.09	-.06
Precise	-.59	.24	-.07*	.00	.01	.00	.01	.06	.00	-.14	.07	-.06
R^2	.01			.00			.01			.01		
F	2.11			.01			4.46**			3.46*		

Note. Clinical was coded as 0 if participants were in the mechanical or hybrid condition, and 1 if participants were in the clinical condition. Hybrid was coded as 0 if participants were in the mechanical or clinical condition, and 1 if participants were in the hybrid condition. Precise was coded as 0 if participants received rounded anchors, and 1 if participants received precise anchors. Interaction terms were dropped due to insignificance.
* $p < .05$. ** $p < .01$. *** $p < .001$

Table 4. *Regressions of Lens Model Parameters on Anchor Features (N = 1045)*

	Judgment Accuracy			Relative Accuracy			Linear Knowledge			Response Consistency		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Clinical	.01	.01	.02	.01	.01	.05	.01	.02	.03	.01	.01	.10
Hybrid	-.02	.01	-.06	-.00	.01	-.03	-.02	.02	-.04	.01	.01	.05
Precise	.01	.01	.04	.00	.01	.02	.01	.02	.04	.01	.01	.08
Clinical x Precise	-.04	.02	-.12*	-.03	.01	-.14*	-.06	.03	-.13*	-.02	.01	-.11*
Hybrid x Precise	-.01	.02	-.03	-.00	.01	-.01	-.02	.03	-.04	-.01	.01	-.06
R^2	.01			.01			.01			.00		
F	2.21			2.49*			2.29*			.93		

Note. Clinical was coded as 0 if participants were in the mechanical or hybrid condition, and 1 if participants were in the clinical condition. Hybrid was coded as 0 if participants were in the mechanical or clinical condition, and 1 if participants were in the hybrid condition. Precise was coded as 0 if participants received rounded anchors, and 1 if participants received precise anchors.
* $p < .05$

Table 5. *Regressions of Lens Model Parameters on Anchor Source by Anchor Precision (N = 1045)*

	Judgment Accuracy						Relative Accuracy					
	Rounded			Precise			Rounded			Precise		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Clinical	.01	.01	.02	-.04	.02	-.13*	.01	.01	.05	-.02	.01	-.13*
Hybrid	-.02	.01	-.07	-.03	.02	-.09	-.00	.01	-.03	-.01	.01	-.04
R^2	.01			.01			.00			.01		
F	1.69			3.39*			1.18			3.45*		

	Linear Knowledge						Response Consistency					
	Rounded			Precise			Rounded			Precise		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Clinical	.01	.02	.04	-.05	.02	-.13*	.01	.01	.10*	-.01	.01	-.05
Hybrid	-.02	.02	-.05	-.04	.02	-.09	.01	.01	.05	-.00	.01	-.03
R^2	.01			.01			.01			.00		
F	1.34			3.22*			2.05			.42		

Note. Clinical was coded as 0 if participants were in the mechanical or hybrid condition, and 1 if participants were in the clinical condition. Hybrid was coded as 0 if participants were in the mechanical or clinical condition, and 1 if participants were in the hybrid condition.

* $p < .05$

Table 6. *Descriptive Statistics of the Magnitude of Anchoring and Lens Model Parameters*

	Average Deviation		Anchor-Judgment Correlation		Importance		Use		Accuracy		Relative Accuracy		Linear Knowledge		Response Consistency	
Condition	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Control	17.49	5.66	.68	.27	-	-	-	-	.38	.20	-	-	.71	.33	.89	.13
ML	10.37	3.48	.87	.10	3.68	0.93	3.72	1.07	.42	.13	.03	.08	.76	.18	.94	.08
MH	9.91	3.67	.88	.10	3.74	1.05	3.53	1.14	.43	.13	.03	.08	.78	.16	.95	.06
CL	10.28	3.89	.87	.10	3.45	1.09	3.49	1.24	.42	.13	.04	.09	.78	.17	.95	.04
CH	9.85	4.82	.86	.17	3.40	1.01	3.27	1.27	.39	.14	.01	.09	.73	.22	.94	.09
HL	10.64	3.74	.87	.10	3.53	1.09	3.47	1.31	.40	.13	.02	.08	.75	.16	.94	.06
HH	9.84	4.16	.87	.11	3.54	1.04	3.47	1.26	.40	.14	.02	.09	.74	.19	.94	.06

Note. ML = Mechanical, low precision condition. MH = Mechanical, high precision condition. CL = Clinical, low precision condition. CH = Clinical, high precision condition. HL = Hybrid, low precision condition. HH = Hybrid, high precision condition.

Table 7. *Correlations between the Magnitude of Anchoring and Lens Model Parameters*

Variable	1	2	3	4	5	6	7	8
Magnitude of Anchoring								
1. Average Deviation	-							
2. Anchor-Judgment Correlation	-.85**	-						
3. Importance	-.28**	.21**	-					
4. Use	-.30**	.22**	.64**					
Lens Model Parameters								
5. Accuracy	-.02	.11**	-.05	-.03	-			
6. Relative Accuracy	.02	.02	-.15**	-.15**	.42**	-		
7. Linear knowledge	.04	.06*	-.11**	-.12**	.74**	.26**	-	
8. Response Consistency	-.43**	.61**	-.07*	-.07*	.28**	.37**	.27**	-
M	10.14	.87	3.56	3.49	.41	.03	.76	.94
SD	3.94	.11	1.04	1.22	.14	.09	.18	.07

Note. N = 1045. *M* = Mean. *SD* = Standard Deviation.
* $p < .05$. ** $p < .01$

Table 8. *Polynomial Regression of Average Deviation on Relative Accuracy (N = 1045)*

	Model 1			Model 2		
	B	*SE B*	*β*	*B*	*SE B*	*β*
Linear	.98	1.44	.02	.91	1.34	.02
Quadratic				83.01	6.58	.36
R^2	.00			.03		
F for change in R^2				36.10***		

*** $p < .001$

Table 9. *Correlations of Individual Differences with Dependent Variables*

Individual Differences	Dependent Variables							
	Average Deviation	Anchor-Judgment Correlation	Importance	Use	Judgment Accuracy	Relative Accuracy	Linear Knowledge	Response Consistency
Experience	.12**	-.09**	.10**	.05	-.08**	-.11**	-.08*	-.14**
Cognitive Ability	.02	.00	-.22**	-.22**	.14**	.20**	.15**	.30**
Decision Styles								
Rational	.00	.06	.02	.01	.05	.08*	.05	.14**
Avoidant	.00	-.04	.01	.00	.02	-.03	.00	-.07*
Dependent	-.05	.03	.11**	.13**	-.02	-.01	-.04	-.02
Intuitive	-.02	.01	.20**	.18**	-.02	-.09**	-.07*	-.10**
Spontaneous	.00	-.02	.07*	.08**	-.04	-.09**	-.06*	-.14**
Resistance								
Unfairness	.00	-.02	-.03	-.07*	-.01	.00	-.01	-.04
Unusefulness	.01	.00	.00	-.02	-.03	-.05	-.01	-.07*

Note. N = 1045 for Average Deviation, Anchor-Judgment Correlation, Importance, Use. N = 1234 for Judgment Accuracy, Relative Accuracy, Linear Knowledge, and Response Consistency. Unfairness = a unit-weighted composite of fairness of algorithms and fairness of standardized tests, higher score indicates perceiving algorithms and standardized tests as less fair. Unusefulness = a unit-weighted composite of usefulness of algorithms and usefulness of standardized tests, higher score indicates perceiving algorithms and standardized tests as less useful.

* $p < .05$. ** $p < .01$

Table 10. *Correlations among Individual Differences*

Variable	1	2	3	4	5	6	7	8	9
1. Experience	-								
2. Cognitive Ability	-.14**	-							
Decision Styles									
3. Rational	.00	.13**	-						
4. Avoidant	-.09**	-.02	-.25**	-					
5. Dependent	-.03	-.06	.06*	.27**	-				
6. Intuitive	.09**	-.15**	-.25**	.10**	.11**	-			
7. Spontaneous	.11**	-.14**	-.55**	.21**	.00	.54	-		
Resistance									
8. Unfairness	-.05	.02	-.10**	.20**	.08**	.08**	.10**	-	
9. Unusefulness	-.05	-.10**	-.09**	.09**	-.04	.17**	.12**	.50**	-
M	1.92	0.06	4.13	2.31	3.10	3.15	2.29	0.00	0.00
SD	0.90	0.82	0.59	0.99	0.79	0.86	0.76	0.98	0.95

Note. N = 1234. *M* = Mean. *SD* = Standard Deviation.
* $p < .05$. ** $p < .01$

Table 11. *Regression Results of Statistically Significant Individual Differences x Source Interactions When Predicting Anchoring*

	B	*SE B*	*β*
Clinical	-.11	.30	-.01
Hybrid	.08	.30	.01
Experience	.69	.27	.12*
Clinical x Cognitive Ability	-.60	.37	-.08
Hybrid x Cognitive Ability	-.83	.38	-.10*
R^2	.01		
F	1.23		

Note. Average Deviation was the dependent variable. Clinical was coded as 0 if participants were in the mechanical or hybrid condition, and 1 if participants were in the clinical condition. Hybrid was coded as 0 if participants were in the mechanical or clinical condition, and 1 if participants were in the hybrid condition.
* $p < .05$.

Table 12. *Regression Results of Statistically Significant Individual Differences x Source Interactions When Predicting Judgment Accuracy*

Individual Differences		*B*	*SE B*	*β*
Experience	Clinical	-.01	.01	-.05
	Hybrid	-.02	.01	-.08*
	Experience	-.03	.01	-.19***
	Clinical x Experience	.03	.01	.10*
	Hybrid x Experience	.02	.01	.09
	R^2	.02		
	F	3.62**		
Cognitive Ability	Clinical	-.02	.01	-.06
	Hybrid	-.02	.01	-.08*
	Experience	.00	.01	.03
	Clinical x Cognitive Ability	.03	.01	.12**
	Hybrid x Cognitive Ability	.01	.01	.02
	R^2	.02		
	F	5.28***		
Unfairness	Clinical	-.02	.01	-.06
	Hybrid	-.02	.01	-.08*
	Experience	.01	.01	.09
	Clinical x Unfairness	-.03	.01	-.12**
	Hybrid x Unfairness	-.01	.01	-.05
	R^2	.01		
	F	2.56*		
Unusefulness	Clinical	-.02	.01	-.06
	Hybrid	-.02	.01	-.08*
	Experience	.01	.01	.10
	Clinical x Unusefulness	-.04	.01	-.15***
	Hybrid x Unusefulness	-.02	.01	-.07
	R^2	.02		
	F	3.55**		

Note. Clinical was coded as 0 if participants were in the mechanical or hybrid condition, and 1 if participants were in the clinical condition. Hybrid was coded as 0 if participants were in the mechanical or clinical condition, and 1 if participants were in the hybrid condition.
* $p < .05$. ** $p < .01$. *** $p < .001$

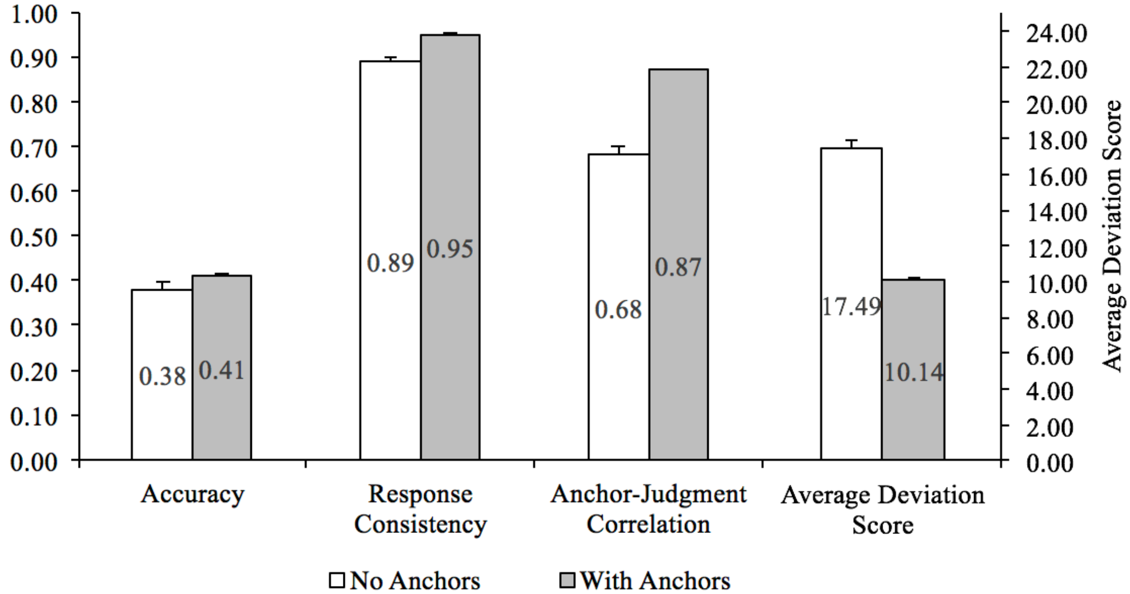


Figure 1. Judgment accuracy, response consistency, and magnitude of anchoring between participants who received anchors and participants who did not receive anchors. Standard errors are represented in the figure by the error bars attached to each column.

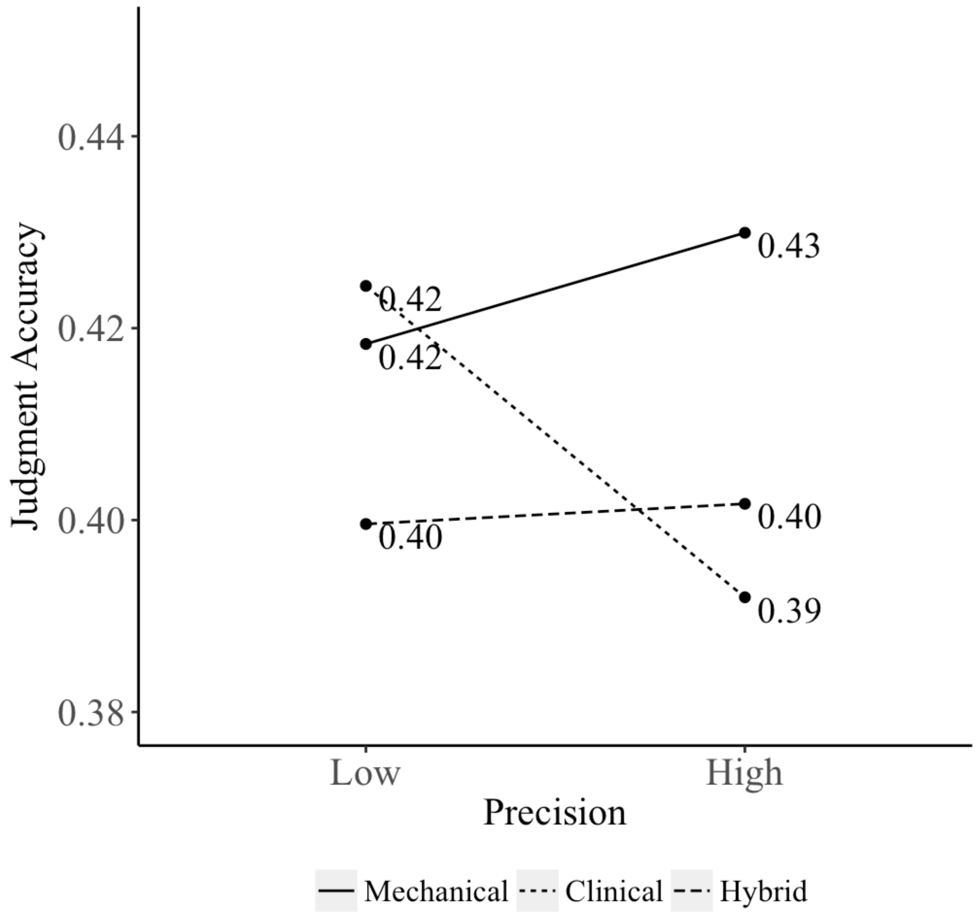


Figure 2. The clinical x precise interaction for judgment accuracy.

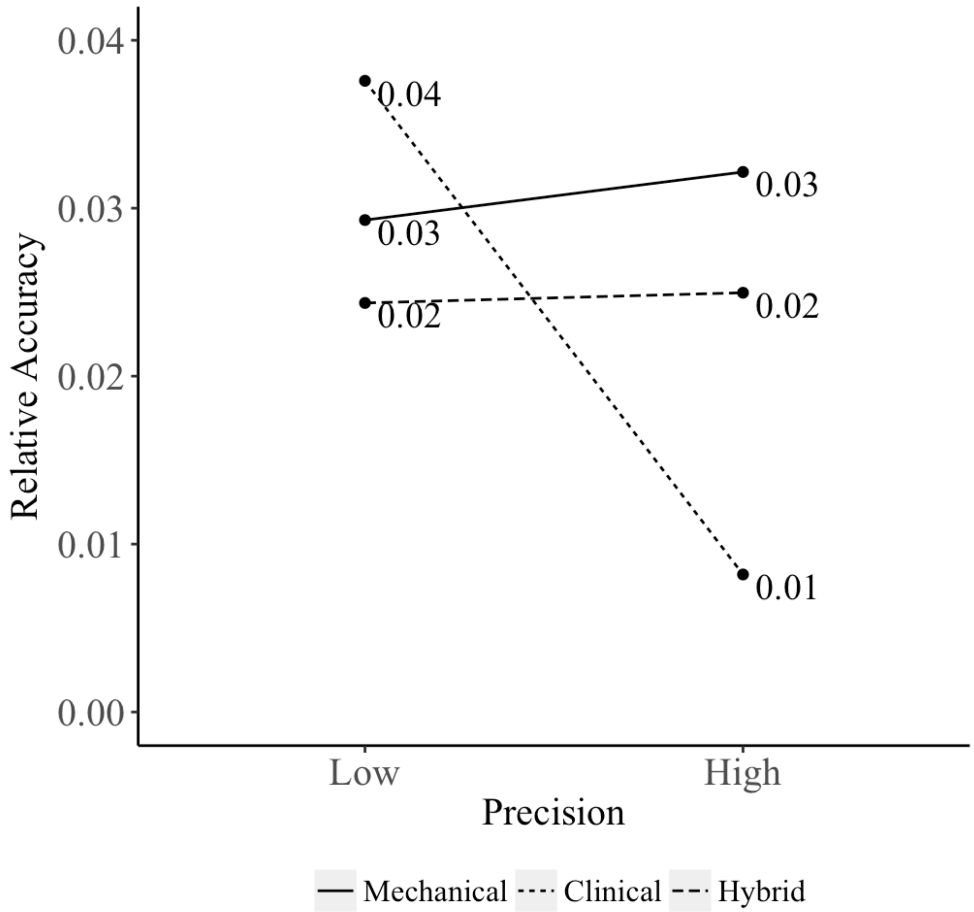


Figure 3. The clinical x precise interaction for relative accuracy.

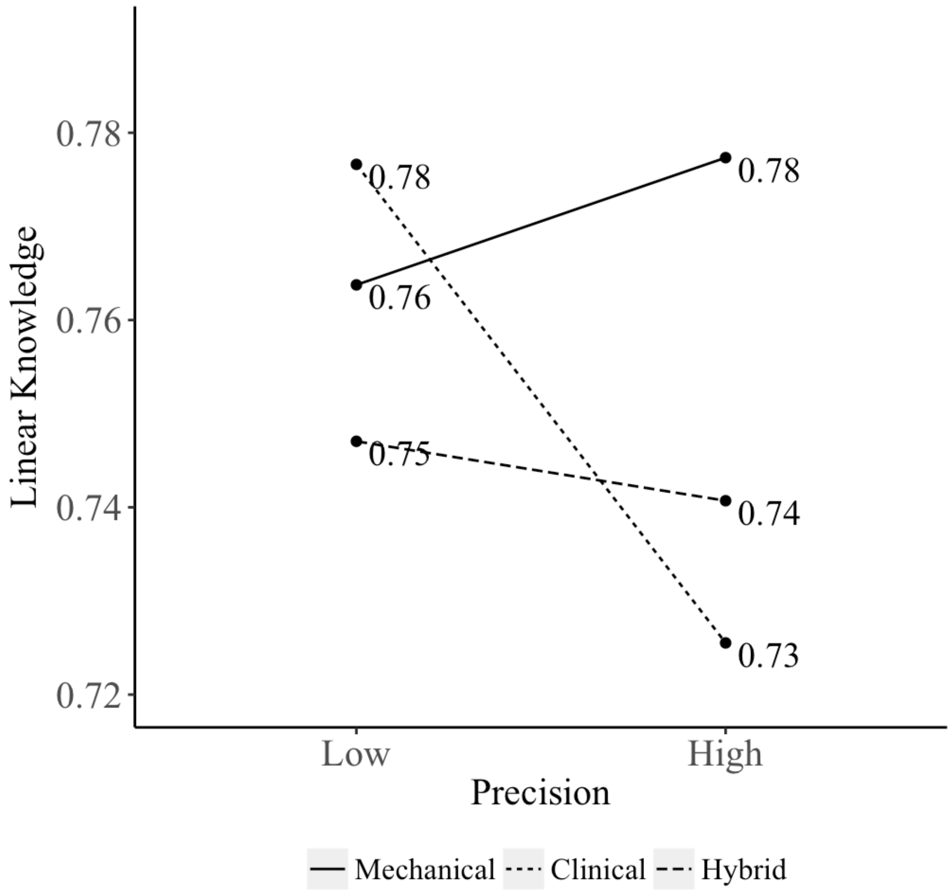


Figure 4. The clinical x precise interaction for linear knowledge.

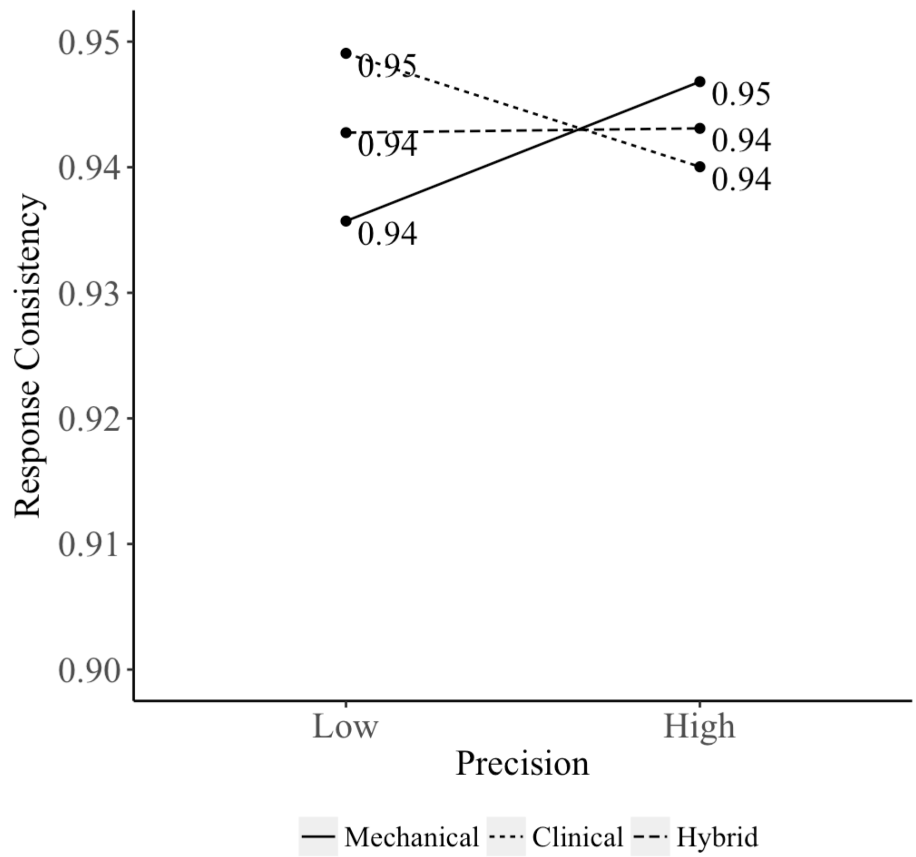


Figure 5. The clinical x precise interaction for response consistency.

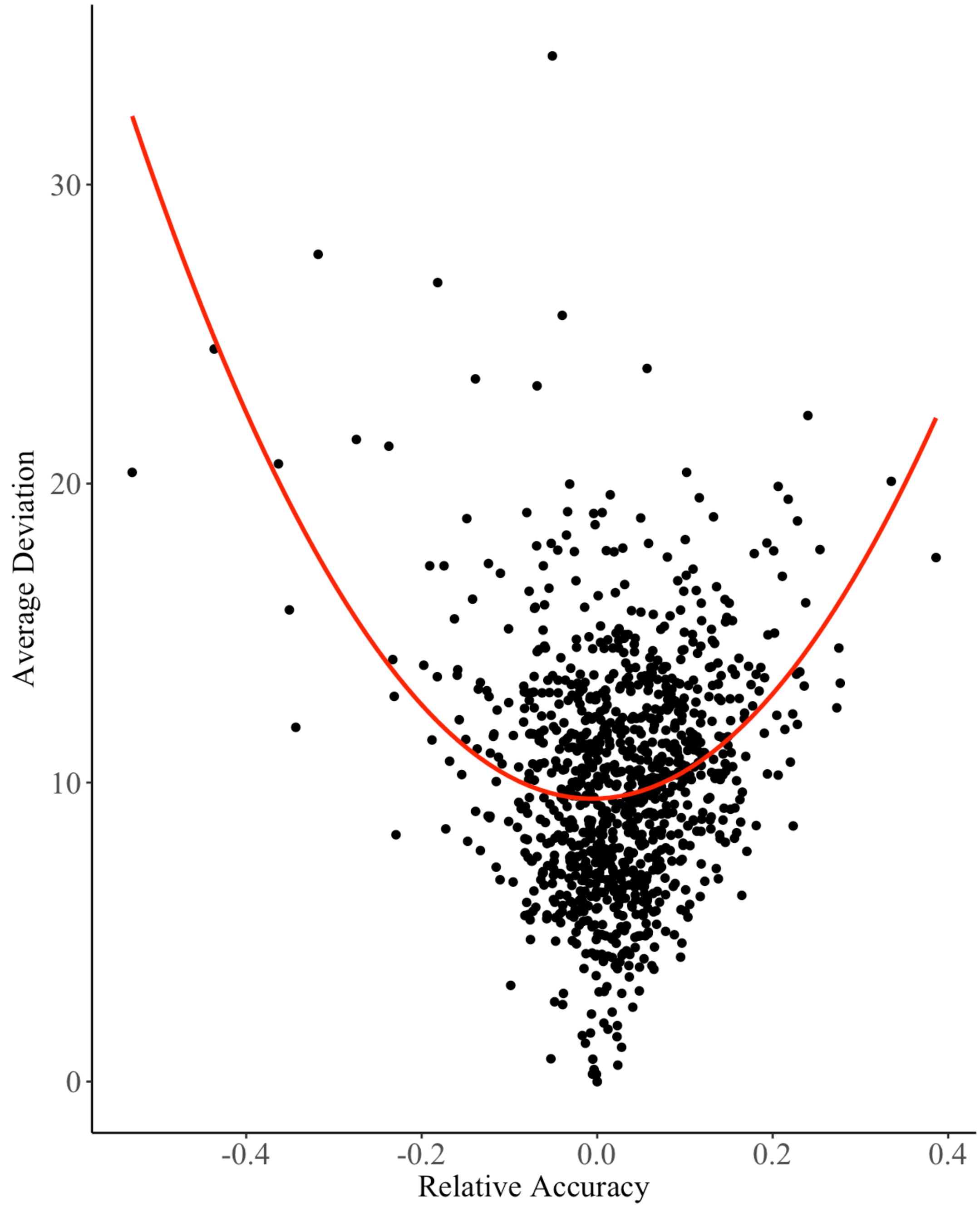


Figure 6. Relationship between relative accuracy and average deviation.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., … Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341–382. https://doi.org/10.1177/0011000005285875

Balzer, W. K., Sulsky, L. M., Hammer, L. B., & Sumner, K. E. (1992). Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes, 53,* 35–54.

Bahník, Š., & Strack, F. (2016). Overlap of accessible information undermines the anchoring effect. *Judgment and Decision Making, 11*, 92–98.

Bavoľár, J. (2017). Experience with the product does not affect the anchoring effect, but the relevance of the anchor increases it. *Ekonomickỳ Časopis (Journal of Economics)*, *3*(65), 282–293.

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, *107*, 66–68. https://doi.org/10.1016/j.econlet.2009.12.028

Bisantz, A. M., Kirlik, A., Gay, P., Phipps, D. A., Walker, N., & Fisk, A. D. (2000). Modeling and analysis of a dynamic judgment task using a lens model approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*, 605-616.

Bodenhausen, G. V., Gabriel, S., & Lineberger, M. (2000). Sadness and the susceptibility to judgmental bias: The case of anchoring. *Psychological Science, 11*, 320–323.

Borman, W. C. (1982). Validity of behavioral assessment for predicting military recruiter performance*. Journal of Applied Psychology, 67*, 3-9.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938-956.

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: University of Chicago Press.

Camerer, D. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can the experts know so much and predict so badly? In K. A. Ericsson and J. Smith (eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 195–217). New York, NY: Cambridge University Press.

Carbone, R., Andersen, A., Corriveau, Y., & Corson, P. P. (1983). Comparing for different time series methods the value of technical expertise individualized analysis, and judgmental adjustment. *Management Science*, *29*, 559–566.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271-280.

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*, 223–242.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes, 79*, 1–39.

Cheek, N. N., & Norem, J. K. (2017). Holistic thinkers anchor less: Exploring the roles of self-construal and thinking styles in anchoring susceptibility. *Personality and Individual Differences*, *115*, 174–176. https://doi.org/10.1016/j.paid.2016.01.034

Cleveland, J. N., & Murphy, K. R. (1992). Analyzing performance appraisal as a goal directed behavior. *Research in Personnel and Human Resource Management, 10*, 121-185.

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

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making, 21*, 241–251.

Crossley, C. D., & Highhouse, S. (2005). Relation of job search and choice process with subsequent satisfaction. *Journal of Economic Psychology, 26,* 255–268.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1774.

Dowd, K. W., Petrocelli, J. V., & Wood, M. T. (2014). Integrating information from multiple sources: A metacognitive account of self-generated and externally provided anchors. *Thinking & Reasoning*, *20*, 315–332. https://doi.org/10.1080/13546783.2013.811442

Englich, B. (2008). When knowledge matters-differential effects of available knowledge in standard and basic anchoring tasks. *European Journal of Social Psychology*, *38*, 896–904. https://doi.org/10.1002/ejsp.479

Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology, 31*, 1535–1551.

Englich, B., & Soder, K. (2009). Moody experts—How mood and expertise influence judgmental anchoring. *Judgment and Decision Making*, *4*, 41-50.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*, 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*, 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Furnham, A., Boo, H. C., & McClelland, A. (2012). Individual differences and the susceptibility to the influence of anchoring cues. *Journal of Individual Differences*, *33*, 89–93. https://doi.org/10.1027/1614-0001/a000076

Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology, 81*, 657–669.

Ganzach, Y., Kluger, A. N., & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology*, *53*, 1-20. doi:10.1111/j.1744-6570.2000.tb00191.x

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, *105*, 387-396.

Glöckner, A., & Englich, B. (2015). When relevance matters: Anchoring effects can be larger for relevant than for irrelevant anchors. *Social Psychology*, *46*, 4–12. https://doi.org/10.1027/1864-9335/a000214

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, *23*, 483-496.

Goldstein, S. G., Deysach, R. E., & Kleinknecht, R. A. (1973). Effect of experience and amount of information on identification of cerebral impairment. *Journal of Consulting and Clinical Psychology*, *41*, 30-34.

Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, *12*, 37-53.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2,* 293–323. doi:10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19–30. doi:10.1037/1040-3590.12.1.19

Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance, 9*, 30–34.

Harding, T. P. (2004). Psychiatric disability and clinical decision making: The impact of judgment error and bias. *Clinical Psychology Review*, *24*, 707–729. https://doi.org/10.1016/j.cpr.2004.06.003

Harris, J. G. (1963). Judgmental versus mathematical prediction: An investigation by analogy of the clinical versus statistical controversy. *Systems Research and Behavioral Science*, *8*, 324–335.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: The Guilford Press.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1,* 333– 342. doi:10.1111/j.1754-9434.2008.00058.x

Highhouse, S., & Kostek, J. A. (2013). Holistic assessment for selection and placement. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology.* (pp. 565–577). Washington: American Psychological Association. https://doi.org/10.1037/14047-031

Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *The Journal of Abnormal and Social Psychology*, *56*, 1-12.

Hough, J. R. & Ogilvie, D. (2005). An empirical test of cognitive style and strategic decision outcomes. *Journal of Management Studies, 42*, 417–48.

Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science*, *19*, 121-127.

Jerez-Fernandez, A., Angulo, A. N., & Oppenheimer, D. M. (2014). Show me the numbers: Precision as a cue to others' confidence. *Psychological Science*, *25*, 633-635.

Johns, G. (1993). Constraints on the adoption of psychology-based personnel practices: Lessons from organizational innovation. *Personnel Psychology, 46*, 569-592.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*, 404-426.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*, 63-82.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*, 1060–1072. https://doi.org/10.1037/a0034156

Leli, D. A., & Filskov, S. B. (1979). The relationship of intelligence with education and occupation as signs of intellectual deterioration. *Journal of Consulting and Clinical Psychology, 47*, 702-707.

Leli, D. A., & Filskov, S. B. (1981). Clinical-actuarial detection and description of brain impairment with the W-B Form I. *Journal of Clinical Psychology, 37*, 623-629.

Leli, D. A., & Filskov, S. B. (1984). Clinical detection of intellectual deterioration associated with brain damage. *Journal of Clinical Psychology, 40*, 1435-1441.

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*, 149–168.

Loschelder, D. D., Friese, M., Schaerer, M., & Galinsky, A. D. (2016). The too-much-precision effect: When and why precise anchors backfire with experts. *Psychological Science*, *27*, 1573–1587.

Loschelder, D. D., Stuppi, J., & Trötschel, R. (2014). "€14,875?!": Precision boosts the anchoring potency of first offers. *Social Psychological and Personality Science*, *5*, 491–499. https://doi.org/10.1177/1948550613499942

Mason, M. F., Lee, A. J., Wiley, E. A., & Ames, D. R. (2013). Precise offers are potent anchors: Conciliatory counteroffers and attributions of knowledge in negotiations. *Journal of Experimental Social Psychology*, *49*, 759–763. https://doi.org/10.1016/j.jesp.2013.02.012

Mathews, B. P., & Diamantopoulos, A. (1989). Judgemental revision of sales forecasts: A longitudinal extension. *Journal of Forecasting*, *8*, 129–140.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Melton, R. S. (1952). *A comparison of clinical and actuarial methods of prediction with an assessment of the relative accuracy of different clinicians*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes, 122*, 69–79.

Muchinsky, P. M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology*, *57*, 175-209.

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes, 98,* 133–143.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35,* 136–164.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin, 26,* 1142–1150.

Nolan, K. P. (2012). *Basic psychological need fulfillment and user resistance to objective and analytical decision-making practices in employee selection* (Unpublished dissertation). Bowling Green State University, Bowling Green, OH.

Nolan, K. P., Carter, N. T., & Dalal, D. K. (2016). Threat of technological unemployment: Are hiring managers discounted for using standardized employee selection practices? *Personnel Assessment and Decisions*, *2*, 30-47.

Nolan, K. P., & Highhouse, S. (2014). Need for autonomy and resistance to standardized employee selection practices. *Human Performance*, *27*, 328–346. https://doi.org/10.1080/08959285.2014.929691

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*, 84–97.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*, 147–152. https://doi.org/10.1016/j.jebo.2009.04.018

Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs, 76*, 1-27.

Perez, F. I. (1976). Behavioral analysis of clinical judgment. *Perceptual and Motor Skills*, *43*, 711–718.

Peterson, D. K., & Pitz, G. F. (1986). Effect of input from a mechanical model on clinical judgment. *Journal of Applied Psychology*, *71*, 163-167.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Russ, F. A., McNeilly, K. M., & Comer, J. M. (1996). Leadership, decision-making and performance of sales managers: A multi-level approach. *Journal of Personal Selling and Sales Management, 3,* 1–15.

Sackett, P. R., Berry, C. M., Wiemann, S. A., & Laczo, R. M. (2006). Citizenship and counterproductive behavior: Clarifying relations between the two domains. *Human Performance, 19*, 441-464.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, *66*, 178-200.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of con- versation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 123–162). San Diego, CA: Academic Press.

Scott, S. G. & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement, 55,* 818–831.

Shu, S., Kuncel, N. R., & Yu, M. C. (2017). *Expert vs. novice judgment in assessments: An experimental test of consistency and accuracy*. Unpublished manuscript.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*, 99-118.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19.

Sjöberg, L. (2003). Intuitive vs. analytical decision making: Which is preferred? *Scandinavian Journal of Management*, *19*, 17–29.

Sjöberg, A., Sjöberg, S., & Forssén, K. (2006). *Predicting Job Performance, PJP. Manual.* Assessio AB (Stockholm).

Smith, A. R., Windschitl, P. D., & Bruchmann, K. (2013). Knowledge matters: Anchoring effects are moderated by knowledge level. *European Journal of Social Psychology*, *43*, 97–108. https://doi.org/10.1002/ejsp.1921

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*, 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*, 437-446.

Terpstra, D.E., & Rozell, E.J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*, 483-495.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review, 71*, 528–530.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Vrieze, S. I., & Grove, W. M. (2009). Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psychology: Research and Practice*, *40*, 525–531. https://doi.org/10.1037/a0014693

Watley, D. J., & Vance, F. L. (1964). *Clinical versus actuarial prediction of college achievement and leadership activity*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Wegener, D. T., Blankenship, K. L., Detweiler-Bedell, B. T., & Petty, R. E. (2009). *Knowledge and anchor extremity. Raw data*. West Lafayette, IN: Purdue University.

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology*, *20*, 5–16. https://doi.org/10.1016/j.jcps.2009.12.003

Weinhardt, J. M., & Beck, J. W. (2017, April). Developing a measure of resistance to scientific personnel selection methods. In Weinhardt, J. M., & Beck, J. W. (Chairs), *Employee selection decision processes: efficient, fair, and accurate.* Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

Welsh, M. B., Delfabbro, P. H., Burns, N. R., & Begg, S. H. (2014). Individual differences in anchoring: Traits and experience. *Learning and Individual Differences*, *29*, 131–140. https://doi.org/10.1016/j.lindif.2013.01.002

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387-402.

Zhang, Y. C., & Schwarz, N. (2013). The power of precise numbers: A conversational logic analysis. *Journal of Experimental Social Psychology*, *49*, 944–946. https://doi.org/10.1016/j.jesp.2013.04.002

Appendix A

Prior Experience with Employee Selection

Concerning your experience evaluating job candidates and making hiring decisions, how often have you …

	Never	A few times	Somewhat often	Often	Very often
reviewed/evaluated application materials	○	○	○	○	○
conducted unstructured employment interviews	○	○	○	○	○
performed reference checks	○	○	○	○	○
conducted structured employment interviews	○	○	○	○	○
reviewed/evaluated samples of work behavior (e.g., simulations)	○	○	○	○	○
reviewed/evaluated profiles posted on social networking websites (e.g., LinkedIn)	○	○	○	○	○
administered or reviewed the results of personality inventories	○	○	○	○	○
administered or reviewed the results of intelligence tests	○	○	○	○	○
been the person who is ultimately responsible for making a hiring decision	○	○	○	○	○

Nolan et al. (2016)

Appendix B

Cognitive Ability Items

Verbal Reasoning

V1. Please mark the word that does not match the other words:
(1) Sycamore (2) Buckeye (3) Elm (4) Daffodil (5) Hickory (6) Sequoia (7) They all match (8) I don't know

V2. The opposite of a "stubborn" person is a "_____" person.
(1) Flexible (2) Passionate (3) Mediocre (4) Reserved (5) Pigheaded (6) Persistent (7) None of these (8) I don't know

V3. Adam and Melissa went fly-fishing and caught a total of 32 salmon. Melissa caught three times as many salmon as Adam. How many salmon did Adam catch?
(1) 7 (2) 8 (3) 9 (4) 10 (5) 11 (6) 12 (7) None of these (8) I don't know

V4. Zach is taller than Matt and Richard is shorter than Zach. Which of the following statements would be most accurate?
(1) Richard is taller than Matt (2) Richard is shorter than Matt (3) Richard is as tall as Matt (4) It's impossible to tell (5) Richard is taller than Zach (6) Zach is shorter than Matt (7) None of these (8) I don't know

V5. If the day after tomorrow is two days before Thursday then what day is it today?
(1) Friday (2) Monday (3) Wednesday (4) Saturday (5) Tuesday (6) Sunday (7) None of these (8) I don't know

V6. Please mark the word that does not match the other words:
(1) Buenos Aires (2) Melbourne (3) Seattle (4) Cairo (5) Morocco (6) Milan (7) None of these (8) I don't know

V7. The opposite of an "affable" person is a(n) "_____" person.
(1) Angry (2) Sociable (3) Gracious (4) Frustrated (5) Reserved (6) Ungrateful (7) None of these (8) I don't know

V8. Isaac is shorter than George and Phillip is taller than George. Which of the following statements is most accurate?
(1) Phillip is taller than Isaac (2) Phillip is shorter than Isaac (3) Phillip is as tall as Isaac (4) It is impossible to tell (5) Isaac is taller than George (6) George is taller than Phillip (7) None of these (8) I don't know

V9. If the day before yesterday is three days after Saturday then what day is today?
(1) Thursday (2) Saturday (3) Wednesday (4) Friday (5) Sunday (6) Tuesday (7) None of these (8) I don't know

V10. The opposite of an "ambiguous" situation is a(n) "_____" situation.
(1) suspicious (2) vague (3) unequivocal (4) intelligent (5) dubious (6) genuine (7) None of these (8) I don't know

Numerical Reasoning

In the following number series, what number comes next?

1) 124,115,106,97,…

2) 85,73,61,49,…

3) 73,55,37,19,…

4) 115,99,83,67,…

5) 200,166,132,98,…

6) 71,65,72,66,73,67,…

7) 23,37,24,38,25,39,…

8) 34,150,35,160,36,170,…

9) 5,16,38,71,115,…

10) 10,24,39,55,72,…

Appendix C

General Decision-Making Style

The rational style emphasizes a thorough search for and logical evaluation of alternatives.

1. I double-check my information sources to be sure I have the right facts before making decisions.
2. I make decisions in a logical and systematic way.
3. My decision making requires careful thought.
4. When making a decision, I consider various options in terms of a specified goal.
5. I explore all of my options before making a decision.

The avoidant style emphasizes postponing and avoiding decisions.

1. I put off making decisions because thinking about them makes me uneasy.
2. I avoid making important decisions until the pressure is on.
3. I postpone decision making whenever possible.
4. I often procrastinate when it comes to making important decisions.
5. I generally make important decisions at the last minute.

The dependent style emphasizes a search for advice and direction from others.

1. I rarely make important decisions without consulting other people.
2. I use the advice of other people in making my important decisions.
3. I like to have someone steer me in the right direction when I am faced with important decisions.
4. I often need the assistance of other people when making important decisions.
5. If I have the support of others, it is easier for me to make important decisions.

The intuitive style emphasizes a reliance on hunches and feelings.

1. When I make decisions, I tend to rely on my intuition.
2. When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.
3. When making a decision, I trust my inner feelings and reactions.
4. When making decisions, I rely upon my instincts.
5. I generally make decisions that feel right to me.

The spontaneous style emphasizes a sense of immediacy and a desire to get through the decision-making process as soon as possible.

1. When making decisions I do what feels natural at the moment.
2. I generally make snap decisions.
3. I often make impulsive decisions.
4. I often make decisions on the spur of the moment.
5. I make quick decisions.

(Scott & Bruce, 1995)

Appendix D

Resistance to Scientific Personnel Selection Methods

Please answer the questions as honestly as possible. We are interested in your opinions about these issues. Some of the questions ask about your preferences when making hiring decisions. Even if you do not currently make hiring decisions, please imagine how you believe you would feel if you needed to make a hiring decision.

1 = strongly disagree
2 = disagree
3 = neither agree nor disagree
4 = agree
5 = strongly agree

Fairness of Algorithms
Hiring algorithms under-predict job performance for racial minorities
Hiring algorithms discriminate against racial minorities
Hiring algorithms are racially biased
Racial minorities are disadvantaged by hiring algorithms
Using hiring algorithms to make hiring decisions is unfair to racial minorities

Fairness of Standardized Tests
Standardized test scores under-predict job performance for racial minorities
Standardized tests discriminate against racial minorities
Standardized tests are racially biased
Racial minorities are disadvantaged by standardized tests
Using standardized tests to make hiring decisions is unfair to racial minorities

Usefulness of Algorithms
Hiring algorithms are not valid for selecting employees
Hiring algorithms do not have value when selecting employees
Hiring algorithms are unreliable for selecting employees
Hiring algorithms do not provide meaningful information about people
Hiring algorithms are useless for selecting employees

Usefulness of Tests
Standardized test scores are not valid for selecting employees
Standardized test scores do not have value when selecting employees
Standardized tests are unreliable for selecting employees
Standardized tests do not provide meaningful information about people
Standardized tests are useless for selecting employees

Appendix E

Background Information and Validity Check Questions

Instruction: Please read the following information carefully. It describes the decision scenario you will encounter in the next section and the data you will use to make subsequent decisions. It is important that you read it carefully prior to proceeding to the next section.

You will answer a few questions on the next page to test if you have read the background information carefully.

An increasing concern among organizations is to reduce counterproductive work behavior.

Counterproductive work behaviors include (but are not limited to):

- Theft
- Destruction of property
- Unsafe behaviors
- Poor quality work (intentionally slow or sloppy work)
- Turnover
- Lateness
- Absenteeism
- Drug and alcohol use
- Workplace violence
- Sexual harassment

One solution is to estimate individuals' integrity scores based on their personality scores.

- Individuals with **higher** integrity scores are **less likely** to engage in counterproductive work behaviors.

Your task in the next section:

- Evaluate the extent to which an individual is likely to exhibit counterproductive work behaviors based on his or her personality scores on five aspects.
- The data you will use to make predictions are from a real employment data set containing personality scores and integrity scores of real employees at an organization, we are able to estimate the accuracy of your predictions.
    - If your accuracy is in the top 25% of individuals who have participated in the study, you will receive a bonus of 1 dollar.

Here are the definitions for each personality score you will encounter:

Extraversion
The extent to which an individual is sociable, gregarious, assertive, talkative, and active.

Agreeableness
The extent to which an individual is compassionate and cooperative rather than suspicious and antagonistic towards others.

Openness
The extent to which an individual is intellectually curious, creative, imaginative, and prefer novelty and variety.

Emotional Stability
The extent to which an individual is calm and emotionally resilient. Individuals high on this are less likely to experience negative emotions.

Conscientiousness
The extent to which an individual is careful, thorough, responsible, organized, hardworking, and achievement-oriented.

There is scientific evidence that individuals who are **higher** on Agreeableness, Emotional Stability, and, **especially, Conscientiousness** are also **higher** on integrity and therefore are **less likely** to engage in counterproductive work behaviors.

Validity Check Questions

1. Is the following statement TRUE or FALSE?

   Examples of counterproductive work behaviors include: poor quality work, absenteeism, turnover, unsafe behaviors, theft, workplace violence, drug and alcohol use, and sexual harassment.

   ○ True
   ○ False

2. Is an individual with a **higher** integrity score less likely, or more likely, to engage in counterproductive work behaviors?

   ○ Less likely
   ○ More likely

3. Which of the following is a characteristic of individuals who are high on Extraversion?

   ○ Intellectually curious
   ○ Talkative
   ○ Compassionate
   ○ Hardworking
   ○ Calm

4. Which of the following is a characteristic of individuals who are high on Agreeableness?

   ○ Intellectually curious
   ○ Talkative
   ○ Compassionate
   ○ Hardworking
   ○ Calm

5. Which of the following is a characteristic of individuals who are high on Openness?

   ○ Intellectually curious
   ○ Talkative
   ○ Compassionate
   ○ Hardworking
   ○ Calm

6. Which of the following is a characteristic of individuals who are high on Emotional Stability?

   ○ Intellectually curious
   ○ Talkative
   ○ Compassionate
   ○ Hardworking
   ○ Calm

7. Which of the following is a characteristic of individuals who are high on Conscientiousness?

   ○ Intellectually curious
   ○ Talkative
   ○ Compassionate
   ○ Hardworking
   ○ Calm

8. Individuals who are **higher** on which of the following personality aspects are **less likely** to engage in counterproductive work behaviors? (Select all that apply)

   ☐ Extraversion
   ☐ Agreeableness
   ☐ Openness
   ☐ Emotional Stability
   ☐ Conscientiousness

Appendix F

Sample Item (Mechanical, High-Precision Condition)

Your task in this section is to predict the integrity score of an individual based on his or her personality scores.

- ⇨ There is scientific evidence that individuals who are **higher** on Agreeableness, Emotional Stability, and, **especially, Conscientiousness** are also **higher** on integrity.
- ⇨ An individual with a **higher** integrity score is **less likely** to exhibit counterproductive work behaviors.

Click here for definitions for counterproductive work behavior and each personality score
If the link does not work, check your popup blocker

Below are the personality scores for Mark:
(On a scale from 0 to 100)

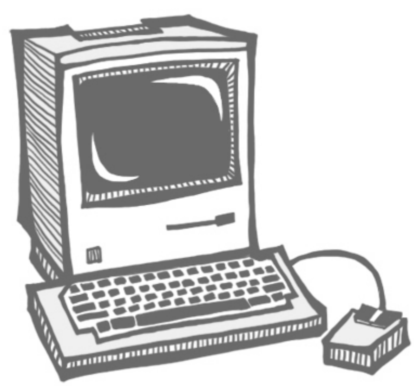
84.9 Extraversion
53.0 Agreeableness
99.6 Openness
29.1 Emotional Stability
15.7 Conscientiousness

Here is how to interpret the personality scores:

A **higher** score means the individual has a **stronger** tendency on that trait.

To guide your judgment, below is a prediction obtained by applying a pre-determined formula to combine the five personality scores.

- There is scientific evidence that an estimate derived through this method is highly accurate, but you are free to make adjustment based on the personality scores.



The **predicted integrity score** based on the formula for Mark is: **64.8** (on a scale from 0 to 100)

Here is how to interpret the integrity score:

A **higher** score means the individual is **less likely** to engage in counterproductive work behaviors.

1. Do you think the **predicted integrity score** of 64.8 based on the formula for Mark is too high, too low, or just appropriate?

○	○	○
Too High	Too Low	Just Appropriate

2. Please provide your judgment of the **integrity score** for Mark based on the personality scores: (on a scale from 0 to 100 where 0 = very likely to exhibit counterproductive work behaviors and 100 = very unlikely to exhibit counterproductive work behaviors)

____

3. Please indicate your **confidence** in your judgment on the following scale:

0 1 2 3 4 5 6 7

Not at all confident Extremely confident