What makes people reject mechanical judgement methods? An investigation of the relationship between individual differences and algorithm use

Sophia Paczulla

Master Thesis – Talent Development and Creativity

s3638324
December 2022
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Prof. Dr. Rob Meijer
Marvin Neumann, MSc.

Abstract

This study aimed to investigate the influence of individual differences on the rejection of mechanical methods in prediction-making. Additionally, it was examined whether a tendency to make use of mechanical methods would result in higher predictive validity. Individual differences that were assessed were the personality factor conscientiousness, its sub facets, experiential thinking, hiring experience, cognitive ability, and advanced professional certification. 308 participants took part in the online survey. The data was analysed by conducting Pearson's correlational analyses and showed that conscientiousness, its sub facets individually, cognitive ability, and advanced professional certification impacted the use of the mechanical method provided in this study. In detail, conscientiousness, its sub-facets, and cognitive ability showed a significant negative correlation with the rejection of the mechanical method and a significant positive correlation with predictive validity. Whereas advanced professional certification showed a significant positive correlation with the rejection of the mechanical method and a significant negative correlation with predictive validity. Only small nonsignificant effects could be found for experiential thinking and hiring experience. The highest correlation was found between the rejection of mechanical methods and predictive validity, which was a significant negative correlation. Future recommendations included replicating this study by using a culturally heterogeneous sample. Lastly, implications entailed tailoring interventions for decision makers more precisely by focusing on people that are less conscientious, score lower in cognitive ability, and do possess advanced professional certification.

Keywords: mechanical judgement, actuarial judgement, individual differences, personality, prediction making

What makes people reject mechanical judgement methods? An investigation of the relationship between individual differences and algorithm use

Making accurate human performance predictions is crucial as this can prevent costly turnover and increase job performance (Hinkin, & Tracey, 2000; Schmidt, & Hunter, 1998). Decision makers typically first collect multiple pieces of information such as test scores and interview impressions, which they then combine to make a decision. Thus, the collection of information can be treated as an antecedent of the information combination (Nikolaou et al., 2015). Decision makers can combine information using one of two approaches, the clinical or the mechanical judgement method (Meehl, 1954; Dawes et al., 1989). In the clinical (or holistic) approach, the decision maker "combines or processes information in his or her head" (Dawes et al., 1989, p. 1668), whereas in the mechanical (or actuarial, statistical) approach, the decision maker is "eliminated and conclusions rest solely on empirically established relations between data and the condition or event of interest" (Dawes et al., 1989, p. 1668). The weights that are assigned to these relations do not have to be solely empirically established. Here, the consistent use of assigned weights is of greater importance. This means that the weights can be chosen to be equal, random, determined based on a meta-analysis or even by the decision maker themselves as long as they are applied consistently (Bobko et al., 2007; Dawes, 1979; Neumann et al., 2021; Yu & Kuncel, 2020).

The Current Problem

The results from various meta-analyses showed that the mechanical method results in higher predictive validity than the clinical method (Dawes et al., 1989; Grove et al., 2000; Kuncel et al., 2013). However, human judges often favour the holistic clinical approach when making predictions (Highhouse, 2008; Portillo, & Mancera-Valencia, 2021; Ryan, & Sackett, 1987; Silzer, & Jeanneret, 2011; Slaughter, & Kausel, 2014; Swets et al., 2000). Many factors have been investigated that explain the use of evidence-based information collection methods

(e.g., structured interviews). However, it is unclear whether they also explain the use of certain combination methods. Since a final decision is typically done after the information combination procedure, this study will focus on the influence of certain individual differences on the use of mechanical judgement methods in information combination. Insights into this topic could be used for educational purposes in the form of interventions that aim to increase the usage of mechanical methods and hence the overall prediction accuracy.

Factors of Influence

Existing research identified several factors that are associated with the rejection of mechanical methods in information collection.

Experiential thinking, experience, and certification

Among other researchers, Lodato et al. (2011) showed that experiential thinking is positively associated with the rejection of mechanical methods. Experiential thinking was described as the tendency to make everyday decisions based on feelings and hunches. Concretely, they suggest that human resource management (HRM) professionals who make everyday decisions based on their intuition also tend to rely on intuition-based approaches when hiring employees. Similarly, Highhouse (2008) proposed that an overreliance on one's intuition impedes the adoption of mechanical methods. When considering these studies, it appears that the rejection of mechanical methods is more likely when the decision maker trusts their intuition, whereas adoption of these methods is more likely when the decision maker bases their judgement on rational thoughts.

Other factors that Lodato et al. (2011) identified that could impair the adoption of mechanical methods are a lack of experience and not possessing an advanced professional certification. The more hiring experience a decision maker has, the less likely they are to reject mechanical methods. They assume that a less experienced decision maker lacks the knowledge about the limitations of human judgement, and thus, sticks with their intuition. In

contrast, Arkes et al. (1986), who investigated conditions in which undergraduates reject a mechanical method such as a decision rule, found that the more experience one has, the less likely one uses a decision rule. They state that experience can lead to overconfidence in one's abilities, which in turn can lead to a feeling that no assistance is needed. Also, Logg et al. (2019) found similar effects. Experienced professionals who make predictions regularly relied less on the help of algorithmic devices. Hence, it is not clear yet how experience and the rejection of mechanical methods relate to each other. Therefore, we will explore this relationship in this study.

Similarly, whether a decision maker holds a Senior Professional in Human Resources (SPHR) certification or not makes a difference in their preference for using either their intuition or a mechanical method. In the case that they do not have such certification, they are more likely to prefer their intuition over mechanical methods (Lodato et al., 2011). It is important to say, however, that Lodato et al. (2011) only investigated peoples' beliefs about relying on feelings in making hiring decisions, whereby this study will assess peoples' actual decision-making behaviour.

Personality

For personality, the factor of conscientiousness was found to positively correlate with the intention to use high-structured interviews (Tsai et al., 2016). Therefore, it is expected that lower levels of conscientiousness predict a lower acceptance of mechanical methods and vice versa. De Vries et al. (2011) suggested that the facets of conscientiousness better predict academic criteria than the general conscientiousness factor. Therefore, we primarily focused on facets rather than the general factor. According to the five-factor inventory by Costa and McCrae (1992), conscientiousness incorporates six facets, which are self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, and cautiousness. There is yet no research available that focuses on the relationship between mechanical judgement

methods and these facets of conscientiousness, or the general factor. However, in line with de Vries et al. (2011), we expect to find higher correlations between these facets and the use of mechanical methods than between the general conscientiousness factor and the use of mechanical methods.

Cognitive ability

De Kock and colleagues (2020) investigated characteristics of a good decision maker, i.e., someone that makes accurate predictions. They reported that cognitive ability was positively related to decision-making accuracy. As an explanation for this finding, they suggested that a decision maker needs to process a lot of information before making a decision, which demands high mental capacities. Since we do not yet know how cognitive ability relates to accuracy when a valid algorithm is available, as well as when participants are informed that using an algorithm will result in better predictions and instructed to make use of it, our aim in this current study is to investigate this relationship. In line with De Kock et al. (2020), we expect to find a negative correlation between cognitive ability and the rejection of mechanical methods under the circumstances that a valid algorithm is included and that participants are informed that using this algorithm will result in better predictions and instructed to make use of it.

Constructs of Interest

According to the research findings, the constructs of (1) experiential thinking (Lodato et al., 2011), (2) (work) experience (Lodato et al., 2011), (3) advanced professional certification (Lodato et al., 2011), the personality dimension (4) conscientiousness (Tsai et al., 2016), and (5) cognitive ability appear to play a role in the rejection of mechanical methods in information collection. Resulting from these research findings in information collection, the following research question is established: “What individual differences explain the rejection of mechanical methods in the information combination process when

making hiring decisions?". Besides this, we hypothesise the following for the process of information combination:

Correlations with Mean Absolute Deviation

Hypothesis 1a: There is a positive correlation between experiential thinking and the rejection of mechanical methods.

Hypothesis 2a: There is a negative correlation between the possession of advanced professional certification and the rejection of mechanical methods.

Hypothesis 3a: There is a negative correlation between the general personality factor conscientiousness, as well as all facets of conscientiousness (i.e., self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, cautiousness) and the rejection of mechanical methods. We expect, however, that there are larger effect sizes between the facets and the rejection of mechanical methods than between the general conscientiousness factor and the rejection of mechanical methods.

Hypothesis 4a: There is a negative correlation between cognitive ability and the rejection of mechanical methods.

Hypothesis 5: There is a negative correlation between the rejection of mechanical methods and the validity of performance predictions.

Correlations with Predictive Validity

Hypothesis 1b: There is a negative correlation between experiential thinking and the validity of performance predictions.

Hypothesis 2b: There is a positive correlation between the possession of advanced professional certification and the validity of performance predictions.

Hypothesis 3b: There is a positive correlation between the general personality factor conscientiousness, as well as all facets of conscientiousness (i.e., self-efficacy,

orderliness, dutifulness, achievement-striving, self-discipline, cautiousness) and the validity of performance predictions.

Hypothesis 4b: There is a positive correlation between cognitive ability and the validity of performance predictions.

Method

The Ethics Committee of the University of Groningen approved this study (PSY-2122-S-0195).

Participants

The sample size for this study was determined by our available budget and was set to 300 participants. To estimate the minimum effect size that we could detect with 80% power, a sensitivity analysis was run by using G*Power 3.1.9.6 (Faul et al., 2009). Results indicated that the minimum important correlation at a significance criterion of $\alpha = .05$ was $r = .14$ (one-tailed) for Correlation: Point biserial model (t tests). After 300 participants were collected, we stopped the data collection. However, after we checked whether this data fulfilled our requirements for participation, the dataset resulted in less than 300 participants. For collecting the remaining ones, we needed to publish our study once more. Since we collected a convenience sample via the online platform Amazon Mechanical Turk (MTurk), a crowdsourcing marketplace, which required a minimum of ten additional participants to be collected, we ended up with data of 308 instead of 300 participants.

Participants had to be fluent in English and had to make at least one hiring decision per year. Next to this, only the ones that worked in an environment where they had to make decisions, solve problems, staff organisational units, and/or judge the qualities of things, services or people were able to proceed with the survey. Lastly, responses were forced for each part of the survey. In between the survey components, attention checks were included, to make sure that no electronic bot nor any inattentive participants were filling out the survey.

For the same reason, we chose to only include the data of participants with a minimum of 15 years between their indicated age and years of hiring experience. Participants that answered any of the attention checks incorrectly were dismissed from the study and not paid anymore. They were informed about this prior to the start of the survey.

Out of the 308 participants, 171 (55.5%) were female, 136 (44.2%) were male, and one (0.3%) person who preferred not to disclose their gender. Data on the participants' demographics are displayed in Table 1. Participants ranged in age from 24 to 63 years (M = 33.87; SD = 9.52). The majority of the sample had a United States of America nationality (89.3%) and identified with a White or Caucasian ethnicity (82.1%). 170 (55.2%) were not SPHR certified, whereas 138 (44.8%) were certified. Participants ranged in years of hiring experience from one to 30 years (M = 6.78; SD = 5.16) and in the number of hiring decisions per year from one to 300 decisions per year (M = 14.27; SD = 21.91).

Table 1

Demographic Data of the Participants (N = 308)

	N = 308
Demographics	*n (%)*
Gender	
Women	171 (55.5)
Men	136 (44.2)
Prefer not to say	1 (0.3)
Nationality	
Albania	1 (0.3)
Armenia	1 (0.3)
Brazil	5 (1.6)
Iceland	1 (0.3)
India	16 (5.2)
Italy	1 (0.3)
New Zealand	1 (0.3)
United Kingdom	7 (2.3)
United States of America	275 (89.3)
Ethnicity	
Asian or Pacific Islander	19 (6.2)
Black or African-American	15 (4.9)
Hispanic or Latino	9 (2.9)
Native-American or Alaskan	9 (2.9)
White or Caucasian	253 (82.1)

Multiracial or Biracial	3 (1)
SPHR certified	
No	170 (55.2)
Yes	138 (44.8)

Design and Procedure

This study used a cross-sectional, correlational design. The online survey was designed with Qualtrics software, a web-based survey tool, distributed via MTurk and it entailed nine sections. The first section was a screener that filtered potential participants based on their work activities. The second section informed participants about the purpose of the study, its procedure, potential participation risks, and informed on participants' data anonymity, data confidentiality and rights. They then were asked for their consent. Thereafter, in the third section, participants were asked to fill in questions concerning the six conscientiousness facets. In the fourth section, participants' levels of experiential thinking were measured.

Then, in the sixth section, they were asked to make 40 performance predictions based on archival applicant data of an airline company (N = 236) originally used in Kausel et al. (2016). The goal of the airline company was to hire ticket agents. Applicants had completed a general mental ability (GMA) test and a conscientiousness questionnaire. Furthermore, a line manager rated applicants' performance in an unstructured interview. Except for the applicants with the lowest possible interview rating, all were hired. Three months after the initial hiring, applicants' overall performance was rated by their supervisors. To ensure that enough participants will participate in our study and that our survey can be completed within an hour, we decided to decrease the number of all applicants to 40. For this, we made sure to select 40 applicants that resemble the original dataset as much as possible, concretely, the absolute differences in correlations between the correlation matrices of the reduced and the full dataset did not differ by more than .015. The R script "Applicant selection.R" that contains the

algorithm that was used to pick the 40 applicants for this study can be found on OSF (https://osf.io/brc9p/?view_only=211357b3a78446078715d47a633d8219). In Appendix C the descriptive statistics and correlations between the variables of the reduced dataset are presented.

Next to applicants' GMA, conscientiousness and interview score, our participants were also given a performance prediction based on a decision rule, which was computed by assigning the following weights to the three assessment scores: applicants' GMA was given 53%, their conscientiousness score was given 28%, and their interview score was given 19%. These regression weights were taken from a meta-analytic correlation matrix presented in Cortina et al. (2000, Table 3). To derive at standardized regression weights from this meta-correlation matrix, we used the *setCor* function from the *psych* package in R (see the R script "Applicant selection.R" on OSF). Participants were informed about these weights. Their task then was to predict the job performance of each of the 40 applicants. To ease participants' interpretation of the predictions, we converted the decision rule prediction to a five-point scale, thus, participants made their performance prediction on a slider between 1 (very poor job performance) and 5 (very good job performance), up to one decimal. Next to this, they were told that they could decide for themselves whether they want to rely on their intuition and expertise to make that performance prediction or to make use of the decision rule. In the latter case, they could simply reproduce the prediction of the decision rule on the slider. However, they were advised to use the decision rule by informing them that research showed its superiority in accuracy compared to intuition and expertise when making predictions. An online application by Failenschmid et al. (2021) was used to check for the predictive validity of the decision rule. The results showed that if participants would strictly use the predictions from the decision rule, their predictive validity would be .35. Applicants were randomized so that each participant received a different applicant order.

In the seventh section, participants had to fill in whether they are a Senior Professional in Human Resources (SPHR) and how often, if ever they read work-related academic journals. After this, in the eighth section, demographics were assessed (i.e., gender, age, nationality, ethnicity) and the number of yearly hiring decisions the professional makes as well as the number of years of hiring experience was asked for. Lastly, in the ninth section, the participants' cognitive ability was measured using two measures of the International Cognitive Ability Resource (ICAR) website, which offers public-domain measurements, the ICAR 9-item letter and number sequence test (ICAR-9) and the ICAR 11-item matrix reasoning test (ICAR-11) (Condon & Revelle, 2014).

The mean study completion time in minutes was M = 38.1 (SD = 22.3). Please note that we identified eight outliers that completed the survey in under ten minutes. However, removing those from the dataset did not change the results, to an extent that would be noteworthy, nor our conclusions, thus, we will report the results for the complete dataset in this study. For completion of the survey, participants were rewarded $6. However, they could earn an additional bonus for making good performance predictions (80¢) and for answering one or more of the cognitive ability items correctly (5¢ for each correctly answered item, thus, up to $1 in total). By this, participants were able to earn up to $7.80 in total.

Measures

Experientiality

The experiential thinking scale was taken from the Rational-Experiential Inventory (REI) by Epstein et al. (1996) and consisted of two subscales, namely, experiential engagement and experiential ability, that were spread over 20 items in total. The scale included positively and negatively worded items. The experientiality scale of the REI had a reliability of α = .82 (experiential engagement (α = .77); experiential ability (α = .59)). Example items looked like "I like to rely on my intuitive impressions." and "I don't think it is

a good idea to rely on one's intuition for important decisions.". Participants had to answer them on a 5-point Likert scale, ranging from "Strongly disagree" to "Strongly agree".

Hiring Experience

Hiring experience was assessed by asking the participants about their years of hiring experience, as well as about the number of hiring decisions they make per year. For this, they had to indicate a number that shows these years or decisions.

Advanced Professional Certification

To assess whether participants were professionally certified, participants had to indicate whether they are Senior Professional in Human Resources (SPHR) certified. They could do so by either clicking "No, I am not." or "Yes, I am.".

Conscientiousness

The conscientiousness facets were measured using Costa and McCrae's (1992) NEO-PI-R facets from the International Personality Item Pool (IPIP) website. The six subscales of conscientiousness were self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, and cautiousness. All scales included five positively and five negatively phrased items, thus ten items in total. The self-efficacy scale ($\alpha = .82$) entailed items like "I handle tasks smoothly." and "I have little to contribute.". The orderliness scale ($\alpha = .77$) incorporated items such as "I like to tidy up." and "I am not bothered by messy people.". Example items of the dutifulness scale ($\alpha = .87$) were "I tell the truth." and "I do the opposite of what is asked.". Items such as "I turn plans into actions" and "I am not highly motivated to succeed." belonged to the achievement-striving scale ($\alpha = .77$). The self-discipline scale ($\alpha = .84$) listed items like "I get chores done right away." and "I postpone decisions.". Example items of the cautiousness scale ($\alpha = .85$) were "I stick to my chosen path." and "I rush into things.". Participants had to answer them on a 5-point Likert scale, ranging from "Strongly disagree" to "Strongly agree".

Cognitive Ability

Cognitive ability was measured by using the ICAR-9 and ICAR-11. The ICAR-9 (α = .66) contained nine questions that looked as follows: "In the following number series, what number comes next? 64, 81, 100, 121, 144, …". The participant then decided between eight answer options. The first six of them provided either a letter or number, the seventh said, "None of these" and the eighth said, "I don't know". The ICAR-11 (α = .62) entailed 11 questions that each showed a matrix consisting of nine elements with one missing element (eight were shown). The participant was then asked to choose from eight answer options. The first six of them provided a letter from A through F with one possible missing matrix element, the seventh said, "None of these", and the eighth said, "I don't know". Condon and Revelle (2014) investigated the reliability and validity of some ICAR scales and reported adequate validity and reliability for the ICAR Sample Test. Although no properties are yet available for the ICAR-9 and ICAR-11, we expected them to have similar validity and reliability estimates.

Rejection of Mechanical Methods

To measure the extent to which the participants reject the use of mechanical methods, our main dependent variable, we assessed their mean deviation from the decision rule. Concretely, as can be seen in the equation below, decision rule deviation was computed as the mean absolute deviation between participants' predicted performance (P) and the decision rule prediction (D) of the 40 predictions (i = 1, …, 40). As a result, scores between 0 and 4 could be obtained. Consequently, the higher the score, the larger the decision rule deviation, and by this, the rejection of mechanical methods.

$$\text{Decision rule deviation} = \frac{\sum_{i=1}^{40} |P_i - D_i|}{40}.$$

Validity of Performance Predictions

Participants' predictive validity was another dependent variable in our study and was operationalized as the correlation coefficient between participants' predictions and applicants' actual performance scores that they were given by their supervisors after three months of hire. Scores between −1 and 1 could be obtained, whereby, scores between −1 and 0 would indicate a negative correlation, meaning, that the higher the participants rated applicants' future performances, the lower applicants' future performances (rated by their supervisors after three months of hiring) were. Whereas scores between 0 and 1 would indicate a positive correlation, meaning that the higher the participants rated applicants' future performances, the higher applicants' future performances were.

Data analysis

The software package IBM SPSS Statistics Version 28 was used for all analyses in this study. First, all participants that did not finish the study for various reasons were removed from the dataset. Afterwards, descriptive statistics were explored, and the statistical distribution of each item was checked. Thereafter, all reversed items were recoded and reliability analyses for the conscientiousness scales, the experientiality scale, the ICAR-9, and the ICAR-11 were conducted. Afterwards, mean variables were computed, i.e., for a mean conscientiousness score, a mean score for all conscientiousness sub scales, a mean experiential thinking score, and a mean rule deviation score. Sum scores were computed for the ICAR-9 and the ICAR-11, as well as a total sum score for both cognitive ability tests together. To summarise participants' predictive validity into one score, the correlation coefficients between participants' prediction scores and applicants' actual performance scores that they were given by their supervisors after three months of hiring were calculated.

After preparing the data for the analyses, one-sided Pearson's correlational analyses were done to check which and to what extent factors significantly correlate with each other. For SPHR certification, conscientiousness facets, as well as the general factor, and cognitive

ability, negative correlations with mean absolute deviation were expected. Whereas for experientiality, a positive correlation with mean absolute deviation was expected. Lastly, for mean absolute deviation, a negative correlation with participants' predictive validity was expected. Afterwards, an exploratory two-sided Pearson's correlational analysis was conducted to investigate the relationship between hiring experience and the rejection of mechanical methods. For all correlational analyses, a correlation coefficient below .30 would be considered weak. Whereby, a value between .30 and .49 would indicate a moderate correlation, and a value of .50 and higher would be considered a strong correlation (Field, 2009). Next to this, a p-value of 0.05 or less was taken as an indicator of a significant correlation.

Results

The first dependent variable of this study, mean absolute deviation from the decision rule, had a mean score of 0.49 (*SD* = 0.40) in our sample. This means that the participants overall did not deviate much from the rule, given that a score of 0 means that they did not deviate at all and a score of 4 means that they highly deviated from the rule. The second dependent variable, participants' predictive validity, had a mean score of .26 (*SD* = 0.15) in this study, which indicates that the participants overall tended to give higher performance prediction scores for applicants that were also rated higher by their supervisors after three months of hiring and vice versa. Participants' predictive validity in this study was thus lower than the predictive validity of the decision rule (.35).

For the independent variables, the experientiality factor had a mean score of 3.24 (*SD* = 0.49) which shows that the participants were moderately intuitive, given that a score of 1 refers to a low score and a score of 5 to a high experientiality score. The general conscientiousness factor as well as all six conscientiousness facets had a mean score that ranged from 3.33 to 3.77 (*SD*s ranged from 0.59 to 0.76) (for more details see Table 2),

which says that the participants in this study were overall moderately conscientious since for all conscientiousness scales a score of 1 indicates a low score and a score of 5 a high conscientiousness score. Cognitive ability had a mean score of 9.52 (*SD* = 4.22). This means that the participants had moderate cognitive abilities when considering that a score of 0 presents a low score and a score of 20 a high cognitive ability score.

Correlations Between the Variables

Correlations with Mean Absolute Deviation

The results of the one-sided Pearson's correlational analyses demonstrated that the correlation between experientiality and mean absolute deviation was nonsignificant and negative (r (308) = –.02, p = .35, 95% Cl [–1.00, 0.07]). Therefore, hypothesis 1a needs to be rejected.

Next to this, there was a significant positive correlation between SPHR certification and mean absolute deviation (r (308) = .18, p < .01, 95% Cl [–1.00, –0.08]). This implies that if a person is SPHR certificated they will score rather high in mean absolute deviation, whereas a person with no SPHR certification will score rather low in mean absolute deviation. Hence, hypothesis 2a needs to be rejected.

The results further demonstrated that there was a significant negative correlation between the general conscientiousness factor and mean absolute deviation (r (308) = –.30, p < .01, 95% Cl [–1.00, –0.21]). This was also the case for all six conscientiousness facets, self-efficacy (r (308) = –.31, p < .01, 95% Cl [–1.00, –0.22]), orderliness (r (308) = –.24, p < .01, 95% Cl [–1.00, –0.15]), dutifulness (r (308) = –.29, p < .01, 95% Cl [–1.00, –0.20]), achievement-striving (r (308) = –.25, p < .01, 95% Cl [–1.00, –0.16]), self-discipline (r (308) = –.24, p < .01, 95% Cl [–1.00, –0.15]), and cautiousness (r (308) = –.29, p < .01, 95% Cl [–1.00, –0.20]). It should be noted, however, that only the correlation between self-efficacy and mean absolute deviation was slightly higher than the correlation between the general

conscientiousness factor and mean absolute deviation. These results show that a person scoring high in conscientiousness, i.e., scoring high in general or on one of the facets, scored rather low in mean absolute deviation and vice versa. Thus, hypothesis 3a can be accepted.

Lastly, there was a significant negative correlation between cognitive ability and mean absolute deviation (r (308) = –.15, p < .01, 95% Cl [–1.00, –0.06]). This means that a person scoring high in cognitive abilities scored rather low in mean absolute deviation and vice versa. Consequently, hypothesis 4a can be accepted.

Mean Absolute Deviation and Predictive Validity

The results of the one-sided Pearson's correlational analyses showed that there is a significant strong negative correlation between mean absolute deviation and participants' predictive validity (*r* (308) = –.72, *p* < –.01, 95% Cl [–1.00, –0.67]). This means that a person scoring high in mean absolute deviation scored rather low in predictive validity and vice versa. Consequently, hypothesis 5 can be accepted.

Correlations with Predictive Validity

The correlation between experientiality and predictive validity was nonsignificant and positive (*r* (308) = .01, *p* = .05, 95% Cl [< –0.01, 1.00]). Thus, hypothesis 1b needs to be rejected.

The results further showed that there was a significant negative correlation between SPHR certification and predictive validity (*r* (308) = –.15, *p* < .01, 95% Cl [0.06, 1.00]). This means that the predictive validity of a SPHR certified person was rather low. Consequently, hypothesis 2b needs to be rejected.

Besides this, there was a significant positive correlation between the general conscientiousness factor and predictive validity (*r* (308) = .32, *p* < .01, 95% Cl [0.24, 1.00]). This was also the case for all six facets, self-efficacy (*r* (308) = .34, *p* < .01, 95% Cl [0.26, 1.00]), orderliness (*r* (308) = .27, *p* < .01, 95% Cl [0.18, 1.00]), dutifulness (*r* (308) = .32, *p* <

.01, 95% Cl [0.23, 1.00]), achievement-striving (r (308) = .31, p < .01, 95% Cl [0.22, 1.00]), self-discipline (r (308) = .28, p < .01, 95% Cl [0.19, 1.00]), and cautiousness (r (308) = .27, p < .01, 95% Cl [0.18, 1.00]). Also here, only the correlation between self-efficacy and predictive validity was slightly higher than the correlation between the general conscientiousness factor and predictive validity. These results imply that a person scoring high in conscientiousness, i.e., scoring high in general or on one of the facets, scored also higher in predictive validity. Consequently, hypothesis 3b can be accepted.

Lastly, there was a nonsignificant positive correlation between cognitive ability and predictive validity (r (308) = .07, p = .12, 95% Cl [–0.03, 1.00]). Therefore, hypothesis 4b can be accepted. A correlation matrix for all variables can be found in Table 2 and an overview of all accepted and rejected hypotheses is shown in Table 4.

Exploratory Analysis

Correlation Between Hiring Experience and Mean Absolute Deviation

There were nonsignificant negative correlations between our two (hiring) experience variables, i.e., years of hiring experience (r (308) = –.03, p = .59, 95% Cl [–0.14, 0.08]) and number of yearly hiring decisions (r (308) = –.08, p = .19, 95% Cl [–0.19, 0.04]), and mean absolute deviation.

Correlation Between Hiring Experience and Predictive Validity

There was a nonsignificant negative correlation between years of hiring experience and predictive validity (r (308) = –.02, p = .75, 95% Cl [–0.13, 0.09]) and a significant positive correlation between number of yearly hiring decisions and predictive validity (r (308) = .12, p = .04, 95% Cl [0.01, 0.23]). Table 3 displays descriptive statistics and the correlations between hiring experience, mean absolute deviation and predictive validity.

Table 2

Means, Standard Deviations, and Pearson Correlations between Independent Variables and Mean Absolute Deviation and Predictive Validity

Variable	*M*	*SD*	Correlations												
			1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Mean Absolute Deviation	0.49	0.40	1.00													
2. Predictive Validity	0.26	0.15	–.72**	1.00												
3. General C	3.60	0.62	–.30**	.32**	1.00											
4. C-Self-efficacy	3.77	0.65	–.31**	.34**	.93**	1.00										
5. C-Orderliness	3.47	0.61	–.24**	.27**	.90**	.78**	1.00									
6. C-Dutifulness	3.71	0.75	–.29**	.32**	.95**	.87**	.82**	1.00								
7. C-Achievement-striving	3.72	0.59	–.25**	.31**	.88**	.83**	.74**	.79**	1.00							
8. C-Self-discipline	3.61	0.70	–.24**	.28**	.93**	.82**	.80**	.84**	.79**	1.00						
9. C-Cautiousness	3.33	0.76	–.29**	.27**	.92**	.79**	.81**	.87**	.70**	.83**	1.00					
10. Experientiality	3.24	0.49	–.02	.01	.32**	.32**	.30**	.28**	.34**	.31**	.23**	1.00				
11. Cognitive ability	9.52	4.22	–.15**	.07	.07	.04	.10*	.09*	.02	.02	.11*	–.06	1.00			
12. Years hiring experience	6.78	5.16	–.03	–.02	–.02	–.03	–.03	–.01	–.01	–.03	–.01	–.16**	–.08	1.00		
13. Number yearly hiring decisions	14.27	21.91	–.08	.12	.06	.09	.05	.06	.06	.07	< –.01	.03	< .01	.10*	1.00	
14. Certification possession			.18**	–.15**	–.37**	–.36**	–.31**	–.35**	–.35**	–.30**	–.36**	–.18**	.13*	–.02	.11*	1.00

Note. ** $p < .01$, * $p < .05$ (one-tailed). C = Conscientiousness. $N = 308$.

Table 3

Means, Standard Deviations, and Pearson Correlations between Hiring Experience and Mean Absolute Deviation and Predictive Validity

Variable	*M*	*SD*	Correlations			
			1.	2.	3.	4.
1. Mean Absolute Deviation	0.49	0.40	1.00			
2. Predictive Validity	0.26	0.15	–.72**	1.00		
3. Years hiring experience	6.78	5.16	–.03	–.02	1.00	
4. Number yearly hiring decisions	14.27	21.91	–.08	.12*	.10	1.00

Note. ** $p < .01$, * $p < .05$ (two-tailed). $N = 308$.

Table 4

Summary of Accepted and Rejected Hypotheses

Hypotheses	Status
H1a: There is a positive correlation between experiential thinking and the rejection of mechanical methods.	Rejected
H2a: There is a negative correlation between the possession of advanced professional certification and the rejection of mechanical methods.	Rejected
H3a: There is a negative correlation between the general personality factor conscientiousness, as well as all facets of conscientiousness (i.e., self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, cautiousness) and the rejection of mechanical methods.	Accepted
H4a: There is a negative correlation between cognitive ability and the rejection of mechanical methods.	Accepted
H5: There is a negative correlation between the rejection of mechanical methods and the validity of performance predictions.	Accepted
H1b: There is a negative correlation between experiential thinking and the validity of performance predictions.	Rejected
H2b: There is a positive correlation between possession of advanced professional certification and the validity of performance predictions.	Rejected
H3b: There is a positive correlation between the general personality factor conscientiousness, as well as all facets of conscientiousness (i.e., self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, cautiousness) and the validity of performance predictions.	Accepted
H4b: There is a positive correlation between cognitive ability and the validity of performance predictions.	Accepted

Discussion

Mechanical methods outperform clinical methods (Dawes et al., 1989; Grove et al., 2000; Kuncel et al., 2013) but are rarely used in practice (Highhouse, 2008; Portillo, & Mancera-Valencia, 2021; Ryan, & Sackett, 1987; Silzer, & Jeanneret, 2011; Slaughter, & Kausel, 2014; Swets et al., 2000). Therefore, the purpose of this study was to investigate the

influence of individual differences on the rejection of mechanical methods. The results revealed that mean absolute deviation is significantly negatively related to predictive validity, the general conscientiousness factor, all six conscientiousness facets, and cognitive ability, and significantly positively related to SPHR certification. Next to this, there were nonsignificant negative correlations between mean absolute deviation and experientiality and the two factors of hiring experience. For predictive validity, significant positive correlations could be found to the general conscientiousness factor, all six conscientiousness facets, and the hiring experience factor of number of yearly hiring decisions, whereas a significant negative correlation could be found towards SPHR certification. Further positive but nonsignificant correlations were shown towards experientiality and cognitive ability, and for the other hiring experience factor, i.e., years of hiring experience, a nonsignificant negative correlation was found. Consequently, five of our nine hypotheses were accepted. Overall, the highest significant effect size was found between mean absolute deviation and predictive validity. Whereby the smallest significant effect sizes could be seen between cognitive ability and mean absolute deviation, as well as number of yearly hiring decisions and predictive validity.

The findings suggest that decision makers who score higher on factors such as conscientiousness and cognitive ability may be less likely to reject mechanical methods. Additionally, decision makers that do possess a SPHR certification will be more likely to reject mechanical methods. The finding of Lodato et al. (2011) that the possession of a SPHR certification can lead to more acceptance of mechanical methods does not support this result. A reason for this could be that our study assessed actual decision-making behaviour, whereas the study by Lodato et al. (2011) only assessed participants' beliefs about relying on feelings in making hiring decisions. Thus, as also known with the intention-behaviour gap, it might be that decision makers with a SPHR certification change their minds and rely on their intuition

when asked to make actual performance predictions, although when asked about their beliefs, they would indicate that relying on one's intuition could harm prediction validity.

For conscientiousness, Tsai et al. (2016) showed that higher levels of conscientiousness are associated with the intention to use high structured interviews. Since high structured interviews allow for less autonomy and by this, less personal judgement of the interviewer, this finding appears to support our result. In line with this, El Othman et al. (2020) studied the effects of personality on decision-making styles in Lebanese universities medical students and found that higher levels of conscientiousness were significantly associated with a rational, less spontaneous, and dependent decision-making style. For the participants in our study to score low in the rejection of mechanical methods, and therefore high in the acceptance of it, they had to reproduce the prediction score of the formula as their own. We argue that this process would call for a more dependent decision-making style since they would follow what is presented to them. Consequently, the finding of El Othman et al. (2020) appears to support our result.

Considering the finding for cognitive ability, De Kock et al. (2020) reported that cognitive ability was related to decision-making accuracy and suggested the explanation that a decision maker needs to process a lot of information before they make a decision, which in turn asks for higher mental capacities. Therefore, we expected to find a negative relationship between cognitive ability and the rejection of mechanical methods, which was also the case. It, therefore, appears that cognitive ability plays an important role in the process of mechanical judgement. Overall, it might be said that decision makers high in cognitive ability may favour the mechanical over the clinical approach when making performance predictions under the condition that a valid algorithm is included which they are informed about and instructed to use. Research that could shed some further light on this association is a study by Dilchert et al. (2007) who studied the effect of cognitive ability on counterproductive work

behaviour (CWB) and reported a negative association, meaning, the higher the level of cognitive ability the lower the level of CWB. To explain, participants in our study were informed that using the decision rule typically results in higher prediction accuracy and were instructed to make use of it. Therefore, one could argue that not using the decision rule and consciously risking making worse predictions can be considered counterproductive work behaviour since the consequence of choosing intuition over algorithm might result in costly turnover for the organisation that the decision maker is working for.

Nonsignificant small effect sizes were found for experientiality and hiring experience. Similar to the finding for SPHR certification, a possible reason for the nonsignificant small effect size between experientiality and the rejection of mechanical methods might be that our study assessed participants' actual behaviour, whereas Lodato et al. (2011) only assessed people's beliefs about relying on feelings in making hiring decisions. When considering hiring experience, the results of existing studies contradicted each other. Lodato et al. (2011) stated that the more hiring experience a decision maker has, the less likely they will reject mechanical methods, whereas Arkes et al. (1986) and Logg et al. (2019) concluded that the more hiring experience a decision maker has, the less likely they are to use mechanical methods. This study was not able to shed further light on the direction of this relationship. Here again, a possible reason for this could be the difference in measuring the rejection of mechanical methods between the study by Lodato et al. (2011) and ours. However, another possible reason could also be the study context differences. To be concrete, Arkes et al. (2011) and Logg et al. (2011) did not focus on predicting future job performance as we did, instead, they let their participants make predictions about basketball players' performance and individuals' weight. Thus, it could be that hiring experience just has no or only a small influence on the rejection of mechanical methods in predicting future job performance.

Strengths, Limitations and Recommendations for Future Research

For considering the strengths of this study, it can be said that it was sufficiently powered to detect practically meaningful effect sizes and that seven out of the ten scales in this study had good reliability, i.e., higher than .70. Next to this, this study was, to our knowledge, the first one that investigated the influence of individual differences on the use of mechanical methods by assessing actual behaviour instead of, for example, intention or attitude.

For the limitations of this study, it can be argued that demographic variables such as ethnicity and nationality were not normally distributed. By this, our sample represented to a high extent white U.S. American citizens. It is commonly known that the United States of America, as a state of western culture, mostly portrays an independent culture. Traits that are typically valued within independent cultures, in comparison to interdependent cultures, are autonomy, independence, and assertiveness (Ma & Schoeneman, 1997; Markus & Kitayama, 1991). As Highhouse (2008) suggested, using mechanical combination procedures might decrease the feeling of autonomy over the decision making. There is a chance that the overrepresentation of people from an independent culture in our sample has influenced the results and that they might be different by taking a more heterogeneous sample, representing independent and interdependent cultures more equally. Thus, the results should be considered with caution. We argue that it could be of value to investigate whether the results of this study can still be achieved by using a culturally heterogeneous sample.

Another recommendation for future research is to replicate this study by including a control group which would not receive the information about the advantage of using the decision rule and is not advised to use it. Our study investigated whether there exist individual differences in who follows this advice. However, it could be that decision makers, high in conscientiousness and cognitive ability and that do not possess a SPHR certification, would make less use of the mechanical method when they are not informed about it and

advised to make use of it. There is existing research that investigated whether informing participants about the advantages of an algorithm makes a difference in people's rejection of mechanical methods. For example, Neumann et al. (2022) found that by showing participants an educational video on evidence-based decision-making, they were more likely to make use of a decision rule immediately after the manipulation and that their predictive validity increased by doing so. However, this effect decreased or disappeared a month later. It is therefore recommended to investigate whether decision makers that are high in conscientiousness and cognitive ability and those that do not possess a SPHR certification will also make use of the mechanical method when they are not provided with any information about it nor advised to make use of it. Moreover, like in the study of Neumann et al. (2022), a follow-up prediction task should be included a month later to check whether these decision makers continue to make use of the mechanical method without again being informed about it before. Study outcomes could give valuable insight for designing interventions as it might be necessary to repeatedly remind decision makers about the advantages of mechanical methods to guarantee that they will also make use of it.

Practical Implications

This study was able to find significant correlations between conscientiousness, cognitive ability, the possession of a SPHR certification and the rejection of mechanical methods in prediction making. However, this study found different results for experientiality and hiring experience than Lodato et al. (2011). It is fair to say that a strength of this study was the assessment of real human behaviour instead of, for example, intention or attitude in prediction making such as in Lodato et al (2011). We, therefore, argue that our findings are important for practice since it is more relevant to look at actual behaviour than at what someone believes in or says they will do. Moreover, the significant negative correlation between the rejection of mechanical methods and participants' predictive validity found in

this study provides support for the initial argument that using mechanical judgement methods will result in higher predictive validity compared to clinical judgement methods.

Overall, our findings can be used to design more tailored interventions. For instance, interventions could focus on decision makers that are less conscientious, less self-efficient, score lower in cognitive abilities and do possess a SPHR certification.

Conclusion

This current paper aimed to investigate the relationship between several individual differences and algorithm use. In particular, we studied what individual differences impact the rejection of mechanical methods. Results showed that the personality factor conscientiousness, as well as all of its six facets, cognitive ability, and professional certification affected the non-use of the provided decision-making rule. The cognitive style of experientiality and hiring experience could not be identified as significant determinants for such rejection. We hope that the insights given by this study can be used to tailor interventions for decision makers more precisely by focusing on people that are less conscientious, score lower in cognitive ability, and do possess a SPHR certification. It is recommended that other researchers replicate this study to explore whether similar results can be obtained by using a more culturally heterogeneous sample since the sample of this study was overrepresented by white U.S. Americans. Another recommendation entailed replicating this study by including a control group that does not receive information on mechanical methods to check whether people high in conscientiousness and cognitive ability and those that do not possess a SPHR certification would still make use of the algorithm when not informed about it nor advised to use it.

References

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 37*(1), 93–110. https://doi.org/10.1016/0749-5978(86)90046-4

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores. *Organizational Research Methods*, *10*(4), 689–709. https://doi.org/10.1177/1094428106294734

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

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, *53*(2), 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*(7), 571–582. https://doi.org/10.1037/0003-066x.34.7.571

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical Versus Actuarial Judgment. *Science, 243*(4899), 1668-1674. https://doi.org/10.1126/science.2648573

De Kock, F. S., Lievens, F., & Born, M. P. (2020). The profile of the 'good judge' in HRM: A systematic review and Agenda for Future Research. *Human Resource Management Review, 30*(2), 100667. https://doi.org/10.1016/j.hrmr.2018.09.003

de Vries, A., de Vries, R. E., & Born, M. P. (2011). Broad versus narrow traits: Conscientiousness and honesty–humility as predictors of academic criteria. *European Journal of Personality*, *25*(5), 336–348. https://doi.org/10.1002/per.795

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology*, *92*(3), 616–627. https://doi.org/10.1037/0021-9010.92.3.616

El Othman, R., El Othman, R., Hallit, R., Obeid, S., & Hallit, S. (2020). Personality traits, emotional intelligence and decision-making styles in Lebanese universities medical students. *BMC Psychology*, *8*(1). https://doi.org/10.1186/s40359-020-00406-4

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*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Failenschmid, J. I., Neumann, M., Meijer, R. R. & Niessen, A. S. M. (2021). *The predictive validity of weighted combinations of predictors and criteria.* Shiny app. Available via: https://failenschmid-neumann-meijer-niessen-2021.shinyapps.io/publication/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Field, A. P. (2009). *Discovering Statistics using IBM SPSS Statistics: and sex and drugs and rock n roll*. London: SAGE Publications Ltd.

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*(1), 19-30. https://doi.org/10.1037/1040-3590.12.1.19

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee

Selection. *Industrial and Organizational Psychology, 1*(3), 333-342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Hinkin, T. R., & Tracey, J. B. (2000). The Cost of Turnover. *Cornell Hotel and Restaurant Administration Quarterly, 41*(3), 14-21. https://doi.org/10.1177/001088040004100313

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27–44. https://doi.org/10.1016/j.obhdp.2016.07.005

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*(6), 1060-1072. https://doi.org/10.1037/a0034156

Lodato, M. A., Highhouse, S., & Brooks, M. E. (2011). Predicting professional preferences for intuition-based hiring. *Journal of Managerial Psychology, 26*(5), 352-365. https://doi.org/10.1108/02683941111138985

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Ma, V., & Schoeneman, T. J. (1997). Individualism versus collectivism: A comparison of Kenyan and American self-concepts. *Basic and Applied Social Psychology, 19*(2), 261–273. https://doi.org/10.1207/s15324834basp1902_7

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295x.98.2.224

Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. https://doi.org/10.1037/11281-000

Neumann, M., Niessen, A. S., Tendeiro, J. N., & Meijer, R. R. (2021). The autonomy-validity dilemma in mechanical prediction procedures: The Quest for a compromise. *Journal of Behavioral Decision Making*, *35*(4). https://doi.org/10.1002/bdm.2270

Neumann, M., Hengeveld, M., Niessen, A. S., Tendeiro, J. N., & Meijer, R. R. (2022). Education increases decision-rule use: An investigation of education and incentives to improve decision making. *Journal of Experimental Psychology: Applied*, *28*(1), 166–178. https://doi.org/10.1037/xap0000372

Nikolaou, I., Oostrom, J. K., & Agnelio, P. (2015). *Employee recruitment, selection, and assessment contemporary issues for theory and practice*. London: Psychology Press.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972–987. https://doi.org/10.1037/0022-3514.76.6.972

Portillo, A. E., & Mancera-Valencia, F. J. (2021). The Processes of Recruitment and Selection of Personnel: Between Objectivity and Subjectivity. *International Journal of Social Sciences Perspectives*, *9*(1), 1. https://doi.org/10.33094/7.2017.2021.91.1.8

Ryan, A. M., & Sackett, P. R. (1987). A Survey Of Individual Assessment Practices By I/o Psychologists. *Personnel Psychology*, *40*(3), 455-488. https://doi.org/10.1111/j.1744-6570.1987.tb00610.x

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262-274. https://doi.org/10.1037/0033-2909.124.2.262

Silzer, R., & Jeanneret, R. (2011). Individual Psychological Assessment: A Practice and Science in Search of Common Ground. *Industrial and Organizational Psychology, 4*(3), 270-296. https://doi.org/10.1111/j.1754-9434.2011.01341.x

Slaughter, J. E., & Kausel, E. E. (2014). Employee selection decisions. In S. Highhouse, R. S. Dalal, & E. Salas (Eds.), Judgment and decision making at work (pp. 57–79). Routledge/Taylor & Francis Group.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better Decisions through Science. *Scientific American, 283*(4), 82-87. https://doi.org/10.1038/scientificamerican1000-82

Tsai, W., Chen, F. H., Chen, H., & Tseng, K. (2016). When Will Interviewers Be Willing to Use High-structured Job Interviews? The role of personality. *International Journal of Selection and Assessment, 24*(1), 92-105. https://doi.org/10.1111/ijsa.12133

Yu, M., & Kuncel, N. (2020). Pushing the limits for judgmental consistency: Comparing random weighting schemes with expert judgments. *Personnel Assessment and Decisions, 6*(2). https://doi.org/10.25035/pad.2020.02.002

Appendix A

Informed Consent

Why do I receive this information?

You are being invited to participate in a research study conducted by Prof. Dr. Rob Meijer, Marvin Neumann (PhD candidate), and Sophia Paczulla (Master student). The aim of this research is to assess and better understand decision-, and prediction-making styles.

Do I have to participate in this research?

Participation in the research is voluntary. However, your consent is needed. Therefore, please read this information carefully. Ask all the questions you might have, for example because you do not understand something. Only afterwards, you decide if you want to participate. If you decide not to participate, you do not need to explain why, and there will be no negative consequences for you. You have this right at all times, including after you have consented to participate in the research.

Why this research?

The aim of this research is to assess and better understand decision-, and prediction-making styles.

What do we ask of you during the research?

Your participation will involve answering several survey questions about you as a person and we will ask you to make 40 predictions about human behavior. We would further like to obtain some demographics of you, such as information about your age, gender, and nationality. Please answer each question in an honest manner. The survey study will take around 40 to 55 minutes to complete.

Your compensation for this study is a monetary reward. You will receive $6 for completing the study. In addition to the compensation that you will receive for completing the study, you

can earn an additional bonus for making good predictions and for answering some of the following questions correctly. We will provide you with more information during the study. Note that we only reward participants who filled in the survey seriously. We will ask you some questions in this study to check whether you carefully read the instructions and whether you are attentive. Please note that we will remove you immediately from this study in the case that you will not answer these questions correctly. If you will be removed, you will also not receive any compensation.

What are the consequences of participation?

We believe that there are no risks by participating in this research. However, there is always the risk of breach possible with this kind of online activity. In the case you feel uncomfortable with partaking in this survey, you may end the survey now. If you feel any discomfort during the survey later on, you have the right to discontinue at any time. You may contact the research team in the case of any questions (contact details are given under 'What else do you need to know?').

How will we treat your data?

The main purpose of the data processing is to write a master's thesis. However, in the case this thesis will be graded sufficient, there might be the chance for further publication of this study. Your data will be processed, collected, prepared, and analyzed by the research team only. While doing so, we will do our best to process your data in the most secure way. This online study will be available between July 1st, 2022, and August 1st, 2022. After this date, further access to the study will not be provided. Your anonymized data will be stored for 10 years. You will have the right to access your personal data until September 1st, 2022, and the right to ask for removal of your data until this date. In the case you wish to do so, please contact one of the researchers (contact details are given under 'What else do you need to

know?'). In the case this study will be published, only the anonymized data will be shared and published by the research team.

What else do you need to know?

Do you have questions/concerns about your rights as a research participant or about the conduct of the research? You may also contact the Ethics Committee of the Faculty of Behavioural and Social Sciences of the University of Groningen: ec-bss@rug.nl.

Do you have questions or concerns regarding the handling of your personal data? You may also contact the University of Groningen Data Protection Officer: privacy@rug.nl.

Contact details of the research team:

Prof. Dr. Rob Meijer: r.r.meijer@rug.nl

Marvin Neumann: m.neumann@rug.nl

Sophia Paczulla: s.paczulla@student.rug.nl (preferred contact)

As a research participant, you have the right to a copy of this research information, for this you might make a screenshot now.

Consent

By pressing **"I agree to participate."**, you **confirm** that you have read the above information and voluntarily decide to participate. You further confirm that you possess a sufficient level of English proficiency.

If you do not wish to participate any longer, please click "I do not agree to participate." And you will be removed from this study.

Appendix B

Exclusion criteria and Attention checks

Exclusion criteria 1: At the beginning of the survey, we presented participants with 12 work activities (in random order). These activities were taken from the O*NET (https://www.onetonline.org/find/descriptor/browse/Work_Activities/). Participants had to indicate up to two tasks in which they most frequently engage at work. They were excluded from our study if they did not choose one of the following options: “Making decisions and solving problems”, “Staffing organizational units”, and “Judging the qualities of things, services, or people”.

Exclusion criteria 2: Next to this, participants were excluded when they indicated that they do not make a single hiring decision per year.

Attention check 1: The conscientiousness scale of orderliness included one item that said “Please click ‘Somewhat agree’ here.”. Only people that clicked on “Somewhat agree” then were allowed to proceed with the survey.

Attention check 2: The experientiality scale also included one item that indicated “Please click ‘Somewhat disagree’ here.”. Again, solely the people that did so were allowed to go on with the study.

Attention check 3: Before participants were informed about the prediction task, they were asked to enter their MTurk-ID for us to be able to pay them for their participation later. After this question, there was a statement that said “I have 115 eyes.”. Participants then had to choose between “Yes.” and “No.”. Only participants that clicked on “No.” were allowed to continue with the survey.

Attention check 4: Before the prediction task started, participants were presented with the information that the decision rule included to derive at a prediction score: a general mental ability score, a conscientiousness score, and a hiring interview rating. On the next

slide, participants were given three different answer options, which each listed possible combinations of assessment ratings (1. A general mental ability test score, an agreeableness questionnaire score, and an interview rating; 2. A general mental ability test score, a conscientiousness questionnaire score, and an interview rating; 3. A physical ability test score, a cognitive ability test score, and an interview rating). They then had to choose the correct one. Only people that clicked on "A general mental ability test score, a conscientiousness questionnaire score, and an interview rating." were allowed to proceed with the study.

Attention check 5: In the section where participants had to answer questions about their work experience and certification, there was one statement included that said "I have 17 fingers.". "Yes." and "No." were the given answer categories. Only people that clicked on "No." could go on and complete the remaining part of the survey.

Appendix C

Descriptive Statistics and Correlations between Variables of the Reduced Applicant Dataset.

Variable	*M*	*SD*	Correlations			
			1.	2.	3.	4.
1. Cognitive ability	0.68	0.14	1.00			
2. Conscientiousness	3.92	0.42	.11	1.00		
3. Interview	2.83	0.98	.11	.02	1.00	
4. Job performance	3.15	0.40	.31	.23	.04	1.00

Note. $N = 40$.