Autonomy-Enhanced-Judgement is a Promising Decision-Tool to Improve Hiring Decisions

Examining Believed Stakeholder Perceptions, Predictive Validity and Threat of Technological Unemployment

Isabell Karolina Budzynski

Master Thesis - Talent Development & Creativity

[*s2848295*]
[*11*] [*2022*]
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Prof. Dr. Rob Meijer and Marvin Neumann (MSc.)

Abstract

Robust evidence favours actuarial methods (i.e., using a decision-rule) over holistic methods (i.e., using intuition) in personnel selection. However, hiring decisions are predominantly based on intuition. Reasons for resistance might be that 1) autonomy-restricting nature of decision-rules 2) decision-makers beliefs about stakeholder perceptions when they use decision-rules and 3) the fear of being replaced by an "algorithm". In the current study, we manipulated autonomy in making performance predictions, expecting that autonomy-enhanced judgement would increase use-intentions and predictive validity. 269 Assessment professionals took part in the online experiment. Participants were randomly assigned one condition and made job-performance predictions for 40 applicants. The conditions allowed varying degrees of autonomy: 1) holistic 2) holistic-adjustment 3) designing a decision-rule, and 4) fixed-rule condition with no autonomy. The main finding was that autonomy-enhanced judgement increased predictive validity over holistic judgement. Second, people were aware of how autonomous and competent stakeholders would perceive them (autonomy-enhanced-judgement improved these perceptions). While the relationship with use-intentions needs to be further explored our results imply that autonomy-enhanced judgement is a promising tool to improve the predictive validity of hiring-decision.

Keywords: Judgement and Decision Making, Human Resource Management, Actuarial Decisions, Holistic Decisions, Autonomy, Threat of Technological Unemployment

Organisations seek to hire the best applicants because selecting suitable personnel is crucial for organisational success (Schmitt et al., 1998). Ideally, the application process identifies the person with the highest suitability for the organisation and the highest chance of performing well in the position (Sekiguchi, 2004). Therefore, the application process should assess the knowledge, skills, and abilities required for the job (Callinan & Robertson, 2000). To assess the desired constructs (i.e., critical thinking, emotional intelligence, leadership skills) an adequate collection method needs to be selected (e.g., work-samples, questionnaires; Arthur & Villado, 2008). In addition, how the information is gathered (e.g., structured versus unstructured interview) also matters in predicting an applicant's future job success (Lievens & De Soete, 2012). Lastly, how the information is *combined* and *weighted* is equally important for the predictive validity of hiring decisions (Grove & Meehl, 1996; Kuncel et al., 2013; Meijer et al., 2020). Substantial efforts are often made to choose the right predictors, but as Yu and Kuncel (2020) remark: *"although certain predictors can be highly valid, ultimately the method used to combine predictor information can serve to either maximize or limit the accuracy of the prediction system"* (p. 1).

Generally, there are two ways to combine information to reach a decision: *holistic* and *actuarial* (Kuncel et al., 2013). *Holistic judgement* (also called clinical judgement) relies on the decision-makers "gut feeling" or mental reasoning to reach a decision. The information is intuitively combined and weighed "in the mind" (Kuncel et al., 2013), selecting the person with the best overall impression. In contrast, in *actuarial judgement* (also called statistical- or mechanical judgement), the information is combined using a decision-rule that is consistently applied to all cases to reach a decision (Kuncel et al., 2013). When organisations employ applicants, hiring decisions are predominantly based on intuition (e.g., Simola et al., 2007). However, robust evidence from nearly 80 years of research shows that using a decision-rule result in more valid predictions than using intuition and, consequently, lead to better hiring-decisions (e.g., Dawes et al., 1989; Grove & Meehl, 1996; Kuncel et al., 2013; Sawyer, 1966). Thus, promoting the acceptance of actuarial hiring approaches is one of the main challenges in implementing more valid hiring decisions (Grove & Meehl, 1996; Highhouse, 2008).

What is a Decision-Rule?

A decision-rule can be as simple as assigning scores to various pieces of information and adding them up (e.g., prediction = predictor_1 * weight_1 + predictor_2 * weight_2 + etc., Dawes & Corrigan,1974; Dawes et al., 1989). One example of a simple decision aid is the NIH Stroke Scale to quantify and predict the severity and impairment of a stroke (Schlegel et al., 2003). The severity is determined by assessing multiple items (e.g., level of consciousness, speech, etc.) on a scale between 0 and 4 and adding them up. Decision-rules can also get more complex when predictors are not weighted equally. Predictor weights can be derived from decision makers' models (e.g., Model of Man; Goldberg, 1970), and optimal weights can be estimated from empirically established predictor-criterion relations (i.e., from primary studies and meta-analyses; Kuncel et al., 2013). For example, a hiring professional can derive predictor weights (e.g., GPA, previous experience, motivation letter) based on the literature on the predictive power for job performance (Meijer et al., 2020). The weights of a rule can also be altered to meet specific objectives for the organisation – for instance, a female- or diversity quota – and may include quantifiable subjective measures (e.g., interview impression; Meijer et al., 2020). Actuarial judgment requires that all predictors are measured on the same scale, that the rule is consistently applied to all cases, and that once the decision-rule is applied, the decision cannot be altered (Meijer et al., 2020). Thus, actuarial judgement is not about *which* predictors are used but *how* the information is integrated. For example, in a hiring decision, the individual chosen by the rule would get the job offer. Once the decision-maker retrospectively adjusts the rule-decision, it cannot be considered pure actuarial judgement (Kuncel et al., 2013).

Actuarial Decision Making and Predictive Validity

In a meta-analysis, Kuncel et al. (2013) compared the predictive validity of actuarial versus holistic decisions in selection and admission decisions: Across 17 examined studies, predictive validity was higher when actuarial judgement was used (Kuncel et al., 2013). Specifically for job performance, the increase in predictive validity converted into a population-level improvement of over 50%. The increase in predictive validity has also been shown in a multitude of work and academic measures (e.g., training outcomes, job performance, supervisor ratings, GPA) and occurs regardless of professional expertise, experience, or knowledge about the job and organisation of the decision-maker

(Kuncel et al., 2013). Even when experts had more information than the decision-rule, did not improve their predictive validity (Kuncel et al., 2013). This highlights how intuitive human judgement, particularly *integrating* relevant information, can interfere with good decision-making.

Superior predictive validity of actuarial methods can be explained twofold. First, to make a good decision, the information needs to be weighted accurately: predictors with high predictive validity receive more weight (Dawes, 1979). Kausel et al. (2016) found that hiring experts do not weight information appropriately. They overestimate information with low predictive validity (i.e., unstructured interviews) and underestimate information with higher predictive validity (e.g., general mental ability). Furthermore, hiring professionals also overestimated their ability to make the right hiring decision (Kausel et al., 2016) and construe stories from irrelevant information (Grove & Meehl, 1996). Actuarial judgement can incorporate predictors with actual predictive validity and calculate optimal weights from empirically established relations (e.g., meta-analyses; Kuncel et al., 2013).

Second, to maximise predictive validity, predictors also need to be weighted consistently (e.g., Yu & Kuncel, 2020, Kuncel et al., 2013). According to Kuncel et al. (2013), people appear to be better at identifying and collecting valid information but are less skilled at information combination. Human decision-makers often weigh the predictors (e.g., motivation letter, job-experience, interview) inconsistently across applicants and tend to focus on salient information (e.g., charismatic applicant; Kuncel et al., 2013). Remarkably, models of decision-makers (called “Model of Man”; i.e., the estimated average predictions of the decision-maker) reliably outperform the decision-maker's actual judgments (called “Man”, Brunswick, 1955; Goldberg, 1970). In other words, following a rule based on previous decisions of the decision-maker results in better decisions than the person the behaviour was modelled from. The difference between the “Man” and “Model of Man” is that predictor weights are applied *consistently*. Moreover, even when the weights of a rule are chosen randomly, the decision-rule still outperforms holistic decisions – provided the individual predictors have some predictive validity (Dawes, 1971; Yu & Kuncel, 2020).

Accurate weighing and consistency *both* seem to have their contribution (Kausel et al., 2016; Yu & Kuncel, 2020). However, consistency might be even more important than the optimal weighting of predictors (Dawes, 1979; Yu & Kuncel, 2020). Implementing decision-rules in practice would increase consistency. However, pure actuarial judgement is – so far – unpopular among assessment professionals. This resistance has been termed ‘algorithm aversion” (Dietvorst et al., 2015). The current study aims to develop and test decision-making systems that help to increase consistency while circumventing this resistance. Nevertheless, the question remains *why* despite robust evidence showing that decision-rules result in better hiring decisions, resistance towards actuarial judgement remains (e.g., Highhouse, 2008; Neumann et al., 2021; Rynes, 2012).

The Reasons for Algorithm Aversion

Some reasons for the underutilisation are a lack of information, misinformation, and mistrust about the utility and validity of actuarial methods (Dawes, 1979; Grove & Meehl, 1996; Highhouse;

2008) combined with inaccurate beliefs about the abilities of human decision-makers (e.g., “they improve with experience”; Highhouse, 2008) and ethical concerns (e.g., “decision-rules are dehumanising”; Dawes, 1979) or insecurity about how to design and apply a decision-rule (Meijer et al., 2020). The decision-maker might be unaware of the benefits of actuarial methods or erroneously believe that their holistic decisions (i.e., “I can see the full picture”) are better than decisions made by a decision-rule (Grove & Meehl, 1996; Rynes, 2012). Especially if the decision-rule makes the same mistake as a human decision-maker, people instantly lose confidence in the rule. Even when people see a decision-rule outperform a human decision-maker, people are still sceptical (Dietvorst et.al., 2015).

Some existing educational interventions seek to alleviate algorithm aversion by highlighting actuarial methods' benefits. Those interventions can lead to increased satisfaction, fairness perceptions, and use-intentions of actuarial methods (Eastwood & Luther, 2016) as well as (temporarily) increasing decision-accuracy (Neumann et al., 2021). Educating assessment professionals might be a good first step. Nonetheless, the ample information available on the benefits of actuarial decision-making, imply that resistance stems from other factors as well. Some important are *autonomy*, *stakeholder perceptions* and the *threat of technological unemployment.*

Reasons for Resistance: Autonomy

According to self-determination theory, autonomy (i.e., experiencing your behaviour as volitional), competence (i.e., feeling capable and skilful), and relatedness (i.e., feeling connected to others) are three basic psychological needs that can explain and predict the motivation perform a specific behaviour (Deci & Ryan, 2000). Behavioural contexts differ in the extent to which they fulfil these needs (Deci & Ryan, 1985). People are less likely to perform a behaviour when they believe it will result in neglecting their needs (Deci & Ryan, 2000).

Research suggests that people resist actuarial decision-methods because they restrict the decision-makers autonomy (Deci & Ryan, 2013; Nolan & Highhouse, 2014). Autonomy is the need to feel choice, control, and agency (Deci & Ryan, 2013). People are susceptible to autonomy loss and will take action to regain it (Radel et al., 2011). Holistic decision-making typically allows a sizeable amount of autonomy, whereas using a strict decision-rule (i.e., actuarial judgement) compromises autonomy by reducing the decision-makers impact on the decision process (Dietvorst et al., 2018; Nolan & Highhouse, 2014; Nolan et al., 2016).

Few studies have examined how to enhance autonomy in actuarial decision-making. This study investigates in which autonomy conditions assessment professionals’ use-intentions towards decision-rules increases. We will investigate whether enhancing autonomy in actuarial decision-methods can increase the willingness to use actuarial judgement. While autonomy is enhanced, we will try to conserve the benefits of using a decision-rule over pure holistic decisions. By identifying reasons that sustain the resistance against actuarial methods and by investigating methods to increase acceptance of actuarial decision-making, we hope to contribute to the establishment of an evidence-based hiring approach in practice.

Enhanced Autonomy in Actuarial Decisions

Autonomy can be enhanced in two ways; people can adjust the predictions a decision-rule with their holistic or "expert" judgement or design the decision-rule themselves (i.e., choose the weight of the rule; Neumann et al., 2020). These decision-methods will be called *autonomy-enhanced judgements* (Neumann et al. 2021).

Some studies have investigated how manipulating the degree of autonomy affects the use-intentions of actuarial judgement (e.g., Dietvorst et al., 2018; Neumann et al., 2021; Nolan & Highhouse, 2014). Nolan and Highhouse (2014) found that increasing structure in the hiring procedure decreases people's autonomy perceptions. Nonetheless, Dietvorst et al. (2018) demonstrated that students used a decision-rule more if they could holistically adjust the predictions. Interestingly, the degree of freedom they were given did not matter. Thus, even slight adjustments to the predictions of a decision-rule might enhance people's perceived autonomy and increase their willingness to use actuarial judgement (Dietvorst et al., 2018).

Another option to enhance autonomy is to choose the weights of the rule (Kuncel et al., 2013; Meijer et al., 2020). Nolan and Highhouse (2014) found that people are more willing to use a decision-rule if they can choose the predictor weights rather than using the weights chosen by someone else. Use-intentions were highest, when the organisation prescribed which criteria to consider during the job interviews (meaning less autonomy in the *data-collection*) while the decision-maker decided how the criteria should be weighted (meaning more autonomy in *data-combination*; Nolan & Highhouse, 2014). These varying effects of autonomy on use-intentions could stem from the fact that people want to avoid choosing an entirely unsuitable applicant for the organization (if the decision-maker chose the decision-criteria), while using this "limited freedom" to exert their final expert decisions to earn praise (Nolan & Highhouse, 2014).

In conclusion, autonomy-enhanced-judgement (adjusting decision-rule predictions or choosing the rule weights) increases people's willingness to use actuarial judgement by increasing perceived autonomy (Dietvorst et al., 2018; Nolan & Highhouse, 2014). Our first research question tries to replicate these findings from Nolan and Highhouse (2014) and answer whether autonomy in the decision-method affects use-intentions. We hypothesise that:

Hypothesis 1a: Use-intentions will be higher for holistic judgement (full autonomy) than for actuarial judgment (fixed rule; no autonomy)

Hypothesis 1b: Use-intentions will be higher for autonomy-enhanced-judgement than for actuarial judgment.

Hypothesis 1c (exploratory): Use-intentions will be similar for autonomy-enhanced-judgement and holistic judgement.

The Predictive Validity of Autonomy-Enhanced-Judgement

Adjusting the predictions of a decision-rule holistically has been shown to decrease predictive validity (Dawes, 1971; Dietvorst et al., 2018). Nonetheless, to justify the utility of autonomy-enhanced-judgement, predictive validity must remain superior to pure holistic decisions. There is evidence that holistic adjustment is more valid than holistic judgement (Dietvorst et al., 2018; Neumann et al., 2021). When people see the predictions of a decision-rule they (un-) intentionally adapt their predictions accordingly. Using the predictions of the decision-rule as an anchor increases the consistency of the predictions, which contributes to the validity of the decisions (Dietvorst et al., 2018; Neumann et al., 2021). Thus, adjusting the rule holistically might not hold the same validity as actuarial judgement, but it could improve holistic decisions.

Similarly, there also is evidence that choosing the weights of the rule can be more valid than holistic predictions. However, to increase decision-accuracy, the predictor validities must be known (Neumann et al., 2021). Nevertheless, we expect that hiring experts have some understanding about the validity of different predictors and give more weight to relevant information (Yu & Kuncel, 2020). Furthermore, consistency might be more important for predictive validity than optimal weighting (Dawes, 1979). Even though hiring experts might not weight predictors optimally (Kuncel et al., 2013), consistency is enhanced when using a self-designed rule (Karelaia & Hogarth, 2008) because the same rule is applied to all applicants.

In conclusion, assessment professionals should follow a decision-rule without *any* adjustment to achieve maximum predictive validity. However, if pure actuarial judgement threatens people's sense of autonomy, and is not used in practice, then strict compliance with actuarial methods might not be the first step to changing current hiring practices. Discovering how much autonomy is necessary to increase acceptance of actuarial decision-making, while maintaining superior predictive validity over pure holistic decisions will provide us with a first understanding whether autonomy-enhanced-judgement shows potential to be utilized in practice. Therefore, our second research question seeks to answer how the decision-method, particularly autonomy-enhanced decision-making, affects predictive validity. Based on the findings on judgement consistency (Goldberg, 1970; Kuncel et al., 2013) and assessment professionals' weighting of predictors (i.e., assuming they have some understanding of the validity of different predictors; Yu & Kuncel, 2020), we believe, that the closer a decision-method resembles pure actuarial judgement (i.e., the fewer inconsistencies the decision-maker can generate across decisions) the higher predictive validity will be. We thus hypothesise that:

Hypothesis 2a (replication): Actuarial judgement (fixed rule condition) will have higher predictive validity than holistic judgement.

Hypothesis 2b: Autonomy-enhanced-judgement will have higher predictive validity than holistic judgement.

Hypothesis 2c (exploratory): Self-designing a rule will have higher predictive validity than holistic adjustment.

Reasons for Resistance: Believed Stakeholder Perceptions and the Threat of Technological Unemployment

According to self-determination theory, autonomy, competence, and relatedness can explain and predict the motivation to perform a specific behaviour (Deci & Ryan, 2000). While restricted autonomy might explain resisting actuarial judgements (Nolan & Highhouse, 2014), people's competence beliefs (feeling capable, skilful and accomplished; Deci & Ryan, 2000) also seem to matter (Nolan, 2012). Generally, people perceive actuarial judgement as more neglective of their competence needs than holistic judgement (e.g., Nolan, 2012). Importantly, during hiring decisions, the decision-process comprises multiple stakeholders (e.g., superiors, colleagues, applicants). Consequently, other people's opinions might also influence the decision-makers attitudes towards actuarial judgement (Nolan, 2012; Nolan et al., 2016).

Nolan et al. (2016) found that decision-makers earn less acknowledgement from stakeholders (e.g., superiors or colleagues) if their hiring decisions are based on a decision-rule rather than their expert judgement. When managers use actuarial judgement stakeholders view them as having less control over the decisions in contrast to when managers use holistic decision-making (Nolan et al. 2016). More importantly, decision-makers are also *aware* of this reduced credit from others (through perspective taking; Davis et al., 2004) and consequently, underutilized decision-rules (Nolan et al., 2016). Thus, decision-makers seem to *know* and *care* what stakeholders think and consequently adjust their behavior (e.g., using a specific hiring-method) to uphold favourable opinions about their capability and to maintain their perceived value for their organization (Nolan et al., 2016; Ryan & Ployhart, 2014).

Meehl (1986) argues that using actuarial decision-making reduces people's professional- and expert values that they believe to offer their organisation, because using a decision-rule reduces the extent to which a decision can be accredited to their expertise. The belief of receiving less credit for a decision might result in the fear of being replaced by an algorithm which jeopardises people's sense of value in the workforce (Meehl, 1986). In other words, people fear that a less qualified person or even a computer could take over their job. This perceived replaceability might result in hesitancy to use decision-rules rules to preserve one's status (Nolan et al., 2016). Meehl (1986) calls this reason for resistance the *"threat of technological unemployment"* (i.e., TOTU; p. 347). Thus, to evoke the threat of technological unemployment stakeholder opinions regarding the decision-makers autonomy (and possibly competence; Nolan, 2012) seem to matter (Nolan et al., 2016). Furthermore, the willingness to use a particular method also depends on the fear of being replaced by algorithms (TOTU; Nolan et al., 2016; 2020).

Nolan et al. (2016) use attribution theory's discounting principle to explain how people attain these beliefs. Attribution theory explains how individuals perceive the cause of events (Kelley, 1973). *Locus of causality* is the perceived cause of either internal or external sources (e.g., personal vs situational factors), and *personal control* is the perceived ability to change an outcome (i.e., agency;

Russell, 1982). In personnel selection, holistic judgement would be considered an internal factor (professional expertise), whereas actuarial judgement would be an external factor (decision-tool; Nolan et al. 2016). When there are several causes possible, the discounting principle proposes that external factors (e.g., actuarial judgement) reduce the extent to which an outcome is attributed to internal factors (e.g., professional expertise; Himmelfarb & Anderson, 1975). As actuarial judgement restricts autonomy, thus restricting personal control and deferring the *locus of causality* more externally, decision-makers believe that they will receive less credit for their hiring decisions (i.e., expertise) from stakeholders (Nolan et al., 2016). Nolan et al. (2016) combined *locus of causality* and *personal control* into one autonomy factor because they were highly correlated – we call this believed stakeholder autonomy attributions (BSAA). Nolan et al. (2016) did not assess decision-maker beliefs on how stakeholders would view their competence when using a specific decision-method. However. in a previous experiment, Nolan (2012) measured competence by asking how the decision-maker would feel using a specific decision-method (e.g., effective, capable, useful, skilful, competent, and accomplished) and found that actuarial judgement fulfilled competence needs less than holistic judgement.

Even though there is evidence that 1) stakeholder perceptions (BSAA) influence decision-makers' attitudes towards actuarial judgement and the threat of technological unemployment (Nolan et al., 2016, Nolan et al., 2020) and 2) that autonomy (Nolan & Highhouse, 2014) and competence (Nolan, 2012) need fulfilment affects use-intentions. It has not been investigated whether decision-makers believed stakeholder competence attributions (BSCA) affect TOTU and use-intentions. Furthermore, it remains unknown whether increasing autonomy (through autonomy-enhanced judgement) affects decision-makers beliefs about the credit they receive from stakeholders. In accordance with Nolan et al. (2016), we believe that decision-makers think that they will be viewed as less autonomous and competent by stakeholders if they use a decision-rule compared to using holistic judgement.

Consequently, our third and fourth research questions seek to answer whether believed stakeholder autonomy attributions (BSAA) and believed stakeholder competence attributions (BSCA based on Nolan, 2012) differ when autonomy-enhanced-judgement is used compared to actuarial judgement. We also want to answer whether autonomy-enhanced-judgement can compete with holistic judgement. Consequently, we will investigate whether BSAA and BSCA are similar when using holistic or autonomy-enhanced judgement. We hypothesise that:

Hypothesis 3a: The believed stakeholder autonomy perceptions (BSAA; locus of causality and personal control) will be lower for actuarial judgement than for holistic judgement.

Hypothesis 3b: BSAA will be lower for actuarial judgement than autonomy-enhanced judgement.

Hypothesis 3c (exploratory): BSAA will be similar for holistic judgment to autonomy-enhanced judgment.

Hypothesis 4a: Believed Stakeholder Competence Attributions (BSCA) will be lower for actuarial judgement than for holistic judgement.

Hypothesis 4b: BSCA will be lower for actuarial judgement than autonomy-enhanced judgement.

Hypothesis 4c (exploratory): BSCA will be similar for holistic condition judgement and autonomy-enhanced judgement.

Furthermore, building on Nolan et al. (2016) and (2020), the fifth research question will investigate how the decision-method (particularly autonomy-enhanced decision-making) affects TOTU. Specifically, we hypothesise that:

Hypothesis 5a: TOTU will be higher for actuarial judgment than holistic judgment.

Hypothesis 5b: TOTU will be higher for actuarial judgment than autonomy-enhanced judgment.

Hypothesis 5c (exploratory): TOTU will be similar for holistic and autonomy-enhanced judgment.

The Mediation Model

In sum, previous research suggests that holistically adjusting decisions from a decision-rule or designing the decision-rule oneself results in a greater willingness to use actuarial decision-making (Dietvorst et al., 2018). Furthermore, it has become evident that improving consistency by using (or approaching) actuarial judgement can enhance predictive validity (e.g., Dawes, 1971; Neumann et al., 2021). With this study, we want to shed light on the association between autonomy-enhanced-judgement and use-intentions of actuarial judgement. As previously discussed, the literature lets reasons to believe that these variables are causally linked (e.g., Nolen et al., 2016; Nolan & Highhouse, 2014). Specifically, we think that use-intentions might be influenced *through* beliefs about stakeholder perceptions of decision-makers autonomy (BSAA) and -competence (BSCA; Nolan, 2012) and the threat of technological unemployment (TOTU; Meehl, 1986).

Thus, our final research question will investigate whether BSAA, BSCA, and TOTU are mediators in the relationship between the method of decision-making (autonomy-enhanced judgement, holistic judgement, and no-autonomy judgement) and use-intentions (Figure 1). We are particularly interested in comparing actuarial judgement and autonomy-enhanced-judgement. Based on attribution theory and earlier findings (Nolan et al., 2016; 2020), it can be expected that enhancing decision-makers autonomy (through choosing the weights or holistic adjustment) increases BSAA and BSCA, which decreases the threat of technological unemployment and which in turn increases use-intentions (Figure 1).

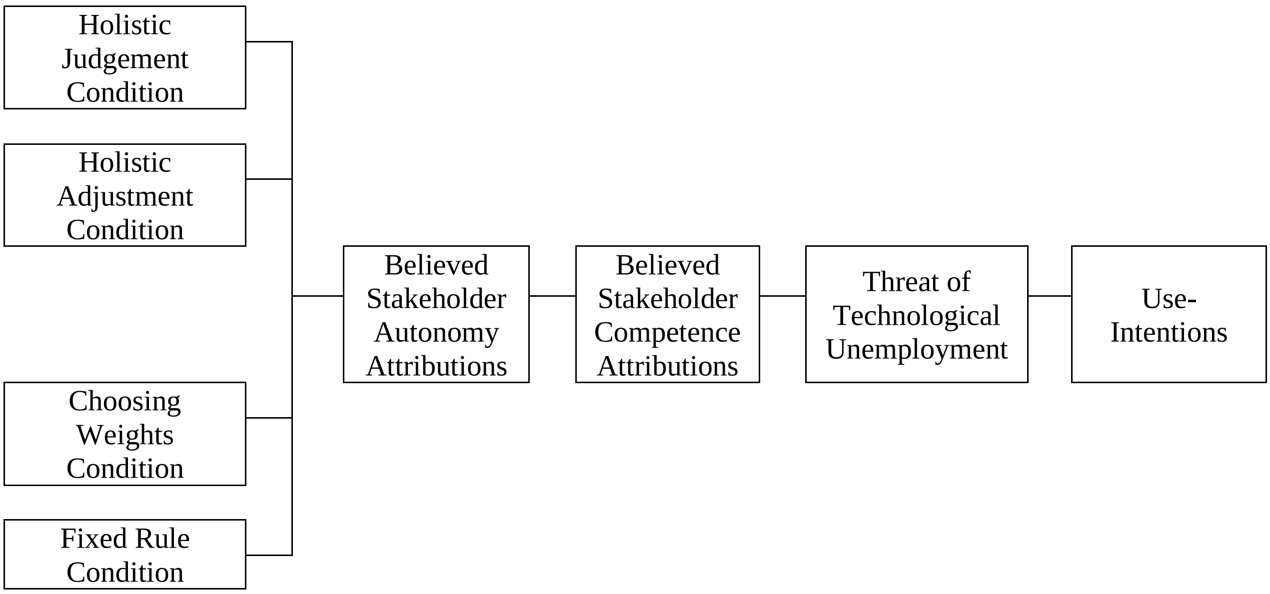
Hypothesis 6: Believed stakeholder autonomy attributions, believed stakeholder competence attributions, and TOTU are mediators in the relationship between the method of decision-making and use-intentions.

The Current Study

The current experiment will employ a one-factor between-subject design with four conditions. The participants will make job performance predictions for 40 applicants based on three pieces of information – a general mental ability test, a standardised personality questionnaire, and an unstructured interview, because of their prevalent use in personnel selection decisions (Farr & Tippins, 2017; Kausel et al. 2016). The participants will be randomly assigned to one condition, which varies in autonomy: In the *holistic condition* (full autonomy) participants will be presented with information and predict future job performance based on expertise and intuition. Further, there are two autonomy-enhanced judgment conditions: 1) In the *holistic-adjustment condition*, the participants will see the performance prediction of a decision-rule as an "anchor point" and decide whether to use it to guide their decision. 2) In the *choosing-weights condition,* participants will create a decision-rule by assigning weights (i.e., importance) to the three pieces of information. Afterwards, the decision-rule will make the predictions for the participants. Lastly, in the *fixed rule condition* (no autonomy), the participants will be presented with the same decision-rule as in the holistic-adjustment condition. They will only see the predictions made by the rule without the possibility to adjust.

Figure 1

Our Conceptual Model



Note. One-factorial between-subjects design with four autonomy conditions. Believed Stakeholder Autonomy Attributions (BSAA; measured by *locus of causality* and *personal control*) and believed stakeholder competence attributions (BSCA) are measured with adapted scales from Nolan (2013). The threat of technological unemployment is measured with a scale by Nolan et al. (2016). Use-intentions are measured with an adapted scale from Nolan and Highhouse (2014).

Method

Participants

The participants were recruited via Amazon Mechanical Turk (MTurk), and the experiment was conducted online using the Qualtrics software (Qualtrics, Provo, UT), Version [09/2]. Prior to the data collection, a G*Power analysis revealed that the between-subject design with four conditions and 80% power requires at least 180 participants when assuming a medium effect size ($\eta p2 = .06$) and $\alpha = .05$ (Cohen, 2013).

A prerequisite for participation was experience in hiring decisions, either currently or in the past. Therefore, various screeners were applied; Firstly, on MTurk, people with the employment status "unemployed" could not access the study. Secondly, the first question on Qualtrics asked participants to indicate up to five tasks in which people most frequently engage at work. If none of the selected tasks related to making hiring decisions (e.g., "Staffing organisational units - recruiting, interviewing, selecting, hiring, and promoting applicants/employees"), the study was automatically discontinued (Appendix B). Thirdly, people were asked how many hiring decisions they undertake in a year. If "0" was selected, the study was also discontinued (Appendix B). To prevent participants from providing socially desirable responses, they were not informed that the survey ended if specific options were chosen. Fourth, we included a comprehension check after explaining the task. If the participant failed to give the correct answer twice, we assumed that the person had not understood the performance prediction task, and the study was discontinued. Lastly, another attention/comprehension check was administered at the end of the survey, in which participants had to indicate their task correctly. Incorrect responses were not included in the analysis.

After the data collection, the participants' response times were checked: finishing the task in under 10 minutes signaled insincerity. Therefore, we excluded those responses from the analysis (pre-registered). The final dataset comprised 269 participants. In the sample, the average time to complete the task was 27 minutes ($SD = 19.61$). Each participant was compensated with 3.65$ for their time. As an incentive to finish the experiment and to increase the "stakes" of the task, participants earned a bonus of up to 5$ for making good performance predictions (Appendix C). The self-reported motivation to obtain the reward – from 1 "extremely unmotivated" to 5 "extremely motivated" – was relatively high in the sample ($M = 4.43$, $SD = 0.64$).

Hiring experience was measured through self-report in years ($M = 5.88$, $SD = 6.07$) and through self-assessment (i.e., "How experienced are you with personnel selection decisions?") on a scale from 1 to 6 ("not experienced" to "extremely experienced", $M = 4.69$, $SD = 0.87$). In terms of occupation, we also measured years of employment ($M = 8.48$, $SD = 6.82$), years in the current role ($M = 5.90$, $SD = 4.84$), employment status (96.3% full-time, 3.7% part-time), occupational title (Table 1) and organisational size (i.e., people working at the organisation; $M = 1426$, $SD = 4971$).

As for the demographics, the final sample consisted of 109 female participants (40.5%), 158 male (58.7%), and 2 other/prefer not to say (0.7%). The mean age was 36.02 (*SD* = 10.17). The predominant nationalities were USA (61.7%) and India (31.6%); the other 6.7% were from the UK, Brazil, and Italy (each, n = 3) and Canada, Colombia, Gambia, Georgia, Germany, Italy, Lithuania, Netherlands, Poland, and Spain (each, n = 1). Ethnicity was composed of 52% White/Caucasian, 31% Asian/Pacific Islander, 10.4% Black/African American, 3.7% Hispanic/Latino and 2.9% other/no indication. 58.7 % of the participants indicated a bachelor's degree as their highest degree of completed education, 31.2% a master's degree, 4,8% high school/secondary education, 3% vocational education, 1.1% doctorate and 1.1% other education.

Table 1

Occupational Titles of the Sample

	N	%
VP/ Director/ Manager	122	45.4
Individual Contributor (e.g., Specialist, Associate, Consultant, etc.)	107	39.8
Other	15	5.6
Entry-Level (Intern, Trainee, Apprentice, Assistant)	13	4.8
CEO/CMO/CHRO or other C-suite job title	12	4.5

Note. Under "other" the participants indicated manager, supervisor, teacher, team leader, operations assistant, and freelancer.

Materials

Stimulus Data

We used real-life applicant stimulus data from Kausel et al. (2016), who collected these pieces of information because of their prevalent use in personnel selection decisions (Farr & Tippins, 2017). Kausel et al. (2016) obtained the data from a selection procedure of an airline company that filled vacancies for a ticket agent. The whole stimulus dataset comprised 236 applicants who were first assessed and scored on three predictors: a general mental ability test score (*GMA*), a standardised personality questionnaire score (*Conscientiousness*), and an unstructured interview score conducted and rated by a line manager (*Interview*). Only the applicants with the lowest interview score were not hired. Three months later, a supervisor rated their overall performance on the job (Kausel et al., 2016). The task of the experiment was to predict applicants' job performance based on the three predictors.

For the current experiment, we sampled 40 applicants out of the 236 for the performance prediction task. More (holistic-) predictions would take longer than reasonable for an online study. To ensure that the correlations in the sample (n = 40) corresponded with the correlations of the whole dataset (N = 236) as closely as possible (by not more than .015), an R-script was created (Appendix A) to sample the applicants.

The three *predictor scores* (GMA, Conscientiousness, Interview) were presented to the participants on the untransformed scales (see Appendix B). The participants were asked to predict the job performance of each applicant: “Based on the information above, how well do you think the applicant will perform in the job assessment three months later in their overall job performance?”. The performance predictions were made on a five-point scale that ranged from “not well at all” to “extremely well” (exact to one decimal). Depending on the condition, the participant either made all 40 performance predictions (holistic; holistic adjustment) or saw ten representative performance predictions of the decision rule (choosing weights; fixed-rule).

Manipulation

Autonomy in making the performance predictions was manipulated. In the *holistic condition,* participants were presented with the predictor scores for all 40 applicants in turn and asked to predict job performance – participants could freely decide how to combine the scores and which information to consider. In the *holistic adjustment condition,* the participants saw the performance prediction of a decision-rule as an “anchor point” and could decide whether to use it to guide their decision. The weights of the decision rule were derived from a meta-analysis by Cortina et al. (2000). The interview ratings in the stimulus data from Kausel et al. (2016) were obtained via an unstructured interview. Therefore, we assumed a level 1 structure (Cortina et al., 2000, p. 339). The percentage weights of the rule were the following: General mental ability test score*53 + Conscientiousness questionnaire score*28 + Interview rating*19 = Decision-rule prediction).

In the *choosing weights, condition,* participants first created a decision-rule by assigning percentage weights to the predictor scores depending on their relative importance. The percentages had to add up to 100 (General mental ability test score*weight_1 + Conscientiousness questionnaire score*weight_2 + Interview rating*weight_3 = Prediction). The stimulus applicant scores (GMA, Conscientiousness, Interview) were standardised and multiplied by the weights participants had indicated. The resulting sum scores were transformed to a 5-point scale, shown to participants because a five-point scale was more comprehensive than the standardised scale. Once the decision-rule was created and shown to the participant, the decision-rule made the job performance predictions. The participant could not make any adjustments to the predictions and watched the rule make 10 decisions to demonstrate how the rule operates. The other 30 predictions were not displayed.

Lastly, participants had no autonomy in the prediction process in the *fixed-rule condition*. The participants were presented with a decision-rule (containing the same weights as in the holistic adjustment condition). The decision-rule then predicted for the participant. The participant could not make any adjustments and just clicked through 10 representative performance predictions.

Measures

The experiment was a one-factor between-subject design with autonomy/method of performance prediction as the four-level between-subject factor (holistic, holistic adjustment, choosing weights, and fixed rule), five dependent variables and a manipulation check:

Perceived autonomy (α = .90) was measured as the manipulation check with an adapted six-item scale (Nolan & Highhouse, 2014). The *perceived competence* (α = .89) was derived from the six-item scale by Nolan (2012). A five-item scale also by Nolan et al. (2016) was used to measure the *threat of technological unemployment* (α = .93). *Use-intentions* (α = .59) were measured with an adapted three-item scale from Nolan & Highhouse (2014). Cronbach's Alpha for the use intention scale improved when only two items were used (1 and 3; α = .85). *Locus of causality* (α = .75) and *personal control* (α = .78) were measured with an adapted three-item scale by Nolan et al. (2016).

The six items from the adapted scales, "locus of causality" and "personal control" (Nolan et al., 2016), were examined with an exploratory principal component analysis. The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = .83 (*"great"*; Field, 2009). Bartlett's test of sphericity ($\chi 2$ (15) = 566.396, $p < 0.001$) was used to obtain eigenvalues for each component in the data. One factor with an eigenvalue > 1 was extracted (namely 3.31; and explained 55.10% of the variance). Table 2 shows the factor loadings. Even though locus of causality and personal control are considered unique dimensions, the principal component analysis supported using a combined measure (in line with Nolan et al., 2016). The factor can be described as *believed stakeholder autonomy attribution* (BSAA*)*. Cronbach's Alpha was higher in the combined scale (α = .83) compared to the individual scales: locus of causality (α = .75) and personal control (α = .78).

For all measures, the mean score across items was calculated (after recoding the reversed item of the use intention scale). For the *predictive validity*, the correlation between the performance predictions (of the participants for the 40 applicants) and the observed performance (of the 40 applicants) was calculated. Thus, there was one correlation per participant that was transformed (Fisher-z transformation) to obtain the analysis-unit.

Table 2

Summary of the Principal Component Analysis for Believed Stakeholder Autonomy Attributions (BSAA)

Item BSAA	Factor Loadings
Think about the other people at your organisation who are familiar with the process used to make this hiring decision. The others would …	
Locus of Causality	
...consider me responsible for the outcome of the decision.	0.74
...think the outcome of this hiring decision reflects on my ability to make hiring decisions.	0.73
...attribute the outcome of this hiring decision to me.	0.74
Personal Control	
...think I had control over how the hiring decision was made.	0.75
...think I had the power to decide which candidate	0.77
...think I was able to change the hiring process as I saw fit.	0.72

Note. N=269, The extraction method was Principal Component Analysis. Because only one component was extracted there was no rotatation. The scales (locus of causality and personal control) were adapted from Nolan et al. (2016).

Procedure

After receiving information about the study, their privacy and data storage, participants gave their active consent to participate. Participants were randomly assigned to one condition without knowing in which condition they were. Participants were given general information about the task, reward scheme, and condition-specific instructions (Appendix C). The comprehension check was administered, after which two practice trials followed. The participants were reminded about the possibility of earning a reward. Afterwards, the actual task followed with either the 40 performance predictions or 10 representative performance predictions (Appendix C) introduced by the following question: *based on the information above* (GMA-, conscientiousness-, interview scores), *how well do you think the applicant will perform in the job assessment three months later in their overall job performance?* In the holistic adjustment and fixed-rule condition, the rule predictions were presented in the following manner: e.g., *"The decision-rule prediction of this applicant's job performance is: 3,7"*. After finishing the task, the participants had to answer a block of questions measuring the dependent variables, demographic questions, and a question about the motivation to obtain the reward (Appendix D).

Data Analyses

The data analysis was done using IBM SPSS Statistics version 27. To answer the research questions, one-way ANOVAs were conducted. The between-subject factor was decision-method, and the dependent variables were BSAA (locus of causality and personal control combined), BSCA, TOTU, use-intentions and predictive validity. Perceived autonomy was included as a manipulation check. If Levene's test for equality of variances was significant, we reported the statistics for equal variances not assumed and Welch's F is reported in the results. The altered degrees of freedom were rounded to the nearest whole number. If the main effect was significant the analyses were followed up with the planned contrast mentioned in the hypotheses: 1) holistic versus fixed-rule 2) fixed-rule versus autonomy-enhanced-judgement (holistic adjustment and choosing weights) and 3) holistic versus autonomy-enhanced-judgement.

To investigate the research question whether BSAA, BSCA and TOTU have a serial indirect effect on the relationship between the method of decision-making and use-intentions, a mediation analysis was done using PROCESS (processv41, Hayes, 2022). The outcome variable in the analysis was *use-intentions,* and *decision-method* was used as the independent variable. Since decision-method is a multi-categorical independent variable, three dummy code variables were created. The fixed-rule condition was used as the reference group because we were interested in how autonomy-enhanced-judgement compared to actuarial judgement. The mediator variables in the analysis were BSAA, BSCA and TOTU. Following Hayes' (2022) PROCESS method (via bootstrapping 5000), we accepted the indirect effects as statistically significant if the 95% bias-corrected CI (lower limit, LLCI; upper limit, ULCI) for the indirect effect (IE) excluded zero.

Results

Manipulation Check

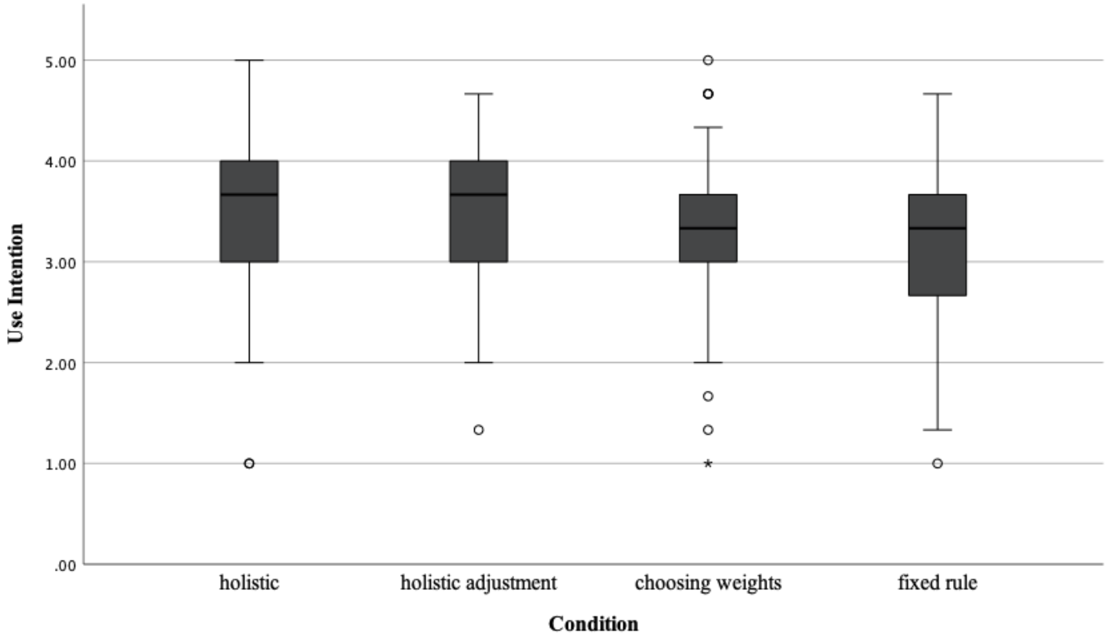
The analysis revealed a significant main effect of decision-method (holistic, holistic adjustment, choosing weights, fixed rule) on autonomy perceptions (Welch(3, 121) = 8.04, $p < .001$, $\omega^2 = .147$, large effect). Autonomy perceptions were lower in the fixed-rule condition than in the holistic condition ($p <.001$). Also, autonomy perceptions were lower in the fixed rule condition than in the choosing weights ($p = .001$; Table 4) and in the holistic adjustment condition ($p < .001$). Autonomy perceptions did not differ in the holistic condition compared to the choosing weights ($p = .341$) and holistic adjustment conditions ($p = .865$). Planned contrasts (unequal variances assumed) also confirmed that autonomy-enhanced-judgement significantly increased autonomy perception compared to actuarial judgement, $t(48.56) = 4.35$, $p < .001$, d = .83 (large effect) and not significantly differed from autonomy perceptions in the holistic condition, $t(116.79) = -1.44$, $p = .153$. Consequently, the planned contrasts also confirmed that actuarial judgement significantly decreased autonomy perceptions compared to holistic judgement $t(53.49) = 4.55$, $p < .001$, d = .91 (large effect). Participants felt less autonomous when using actuarial judgement compared to holistic judgement or autonomy-enhanced-judgement. There was no significant difference in how autonomous participants perceived autonomy-enhanced-judgement compared to holistic judgement.

Use-Intentions

There was no significant main effect of decision-method (holistic, holistic adjustment, choosing weights, fixed rule) on use-intention ($F(3, 265) = 1.29$, $p = .277$). Because the effect was non-significant, no posthoc tests or planned contrasts were conducted. Decision-method did not affect participants' willingness to use a particular method: neither autonomy-enhanced judgement, nor holistic judgement appeared to significantly increase use-intentions over pure actuarial judgement (Table 4). Explorative analyses showed that perceived autonomy significantly correlated with use-intentions ($r = .46$, $p < .01$; Table 3).

Figure 2

Boxplot of Use-Intention per Condition



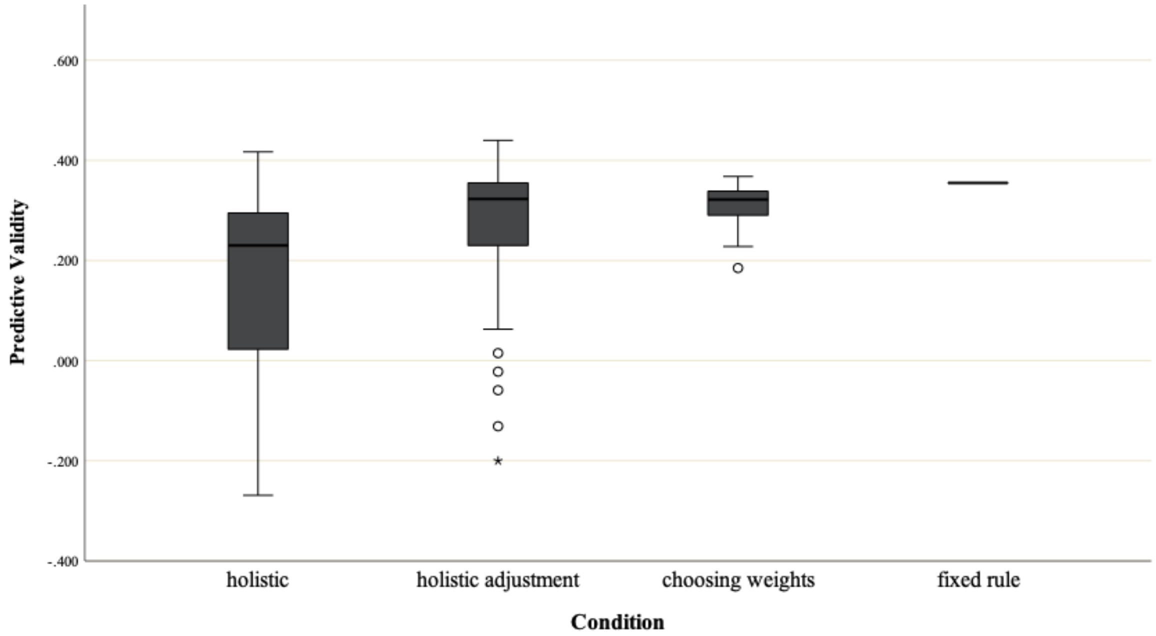
Note. Questions measuring use-intentions can be found in Appendix D.

Predictive Validity

There was a significant main effect of decision-method on predictive validity, (Welch(3, 117) = 69.19, p < .001, ω^2 = .242, large effect). In the fixed rule condition predictive validity was higher than in the holistic condition (p < .001; Table 4; also see Figure 3). Predictive validity was higher in the fixed rule condition than in the choosing weights (p < .001) and in the holistic adjustment condition (p < .001). Predictive validity did not differ in the holistic condition compared to the choosing weights (p = .999) and holistic adjustment conditions (p < .001). Furthermore, predictive validity was higher in the choosing weights condition than in the holistic adjustment condition (p = .034) and higher in the holistic-adjustment condition than in the holistic condition (p = .001). Planned contrasts (unequal variances assumed) confirmed that actuarial judgement (fixed rule) significantly increased predictive validity compared to holistic judgement, t(53.00) = -7.62, p < .001, d = 1.10 (large effect), and to autonomy-enhanced-judgement (holistic adjustment & choosing weights), t(111.38) = -9.34, p < .001, d = .64 (medium effect). Furthermore, the third contrast confirmed that holistic judgement significantly decreased predictive validity compared to autonomy-enhanced-judgement, t(59.89) = 5.10, p < .001, d = .95 (large effect).

Figure 3

Boxplot of Predictive Validity per Condition



Note. Predictive validity was measured as the correlation between performance predictions of the participants (predictor) and the observed performance of the applicants from the stimulus dataset (criterion).

Table 3

Pearson Correlations and Descriptives for Study Variables

	M	*SD*	Perceived Autonomy	BSAA	BSCA	TOTU	Use-Intentions	Predictive Validity
Perceived Autonomy	3.99	.84	-					
BSAA	4.06	.69	.65**	-				
BSCA	4.22	.69	.70**	.65**	-			
TOTU	3.11	1.18	-.04	.01	-.22**	-		
Use-Intentions	3.38	.74	.46**	.33**	.46**	-.23**	-	
Predictive Validity	-	-	-.18**	-.11	-.10	-.22**	-.05	-
Age	36.02	10.17	-.15*	-.08	-.06	-.30**	-.07	.12*

** $p < 0.01$ level; * $p < 0.05$ level; N=269

Table 4

Descriptive Statistics of the Measured Variables per Condition

variable	condition	*n*	*m*	*sd*	min	max	standard error
Perceived Autonomy	holistic	54	4.23	0.52	2.67	5.00	0.07
	holistic adjustment	93	4.15	0.69	1.00	5.00	0.07
	choosing weights	79	4.05	0.68	1.50	5.00	0.08
	fixed rule	43	3.25	1.24	1.00	5.00	0.19
Use-Intentions	holistic	54	3.48	0.81	1.00	5.00	0.11
	holistic adjustment	93	3.42	0.61	1.33	4.67	0.06
	choosing weights	79	3.38	0.73	1.00	5.00	0.08
	fixed rule	43	3.19	0.89	1.00	4.67	0.14
TOTU	holistic	54	3.02	1.34	1.00	4.60	0.18
	holistic adjustment	93	3.31	1.12	1.00	4.80	0.12
	choosing weights	79	3.00	1.16	1.00	5.00	0.13
	fixed rule	43	2.99	1.11	1.00	5.00	0.17
BSAA	holistic	54	4.15	0.58	2.17	5.00	0.08
	holistic adjustment	93	4.13	0.55	2.17	5.00	0.06
	choosing weights	79	4.16	0.58	2.00	5.00	0.06
	fixed rule	43	3.63	1.03	1.00	4.67	0.16
BSCA	holistic	54	4.39	0.54	2.50	5.00	0.07
	holistic adjustment	93	4.27	0.6	1.67	5.00	0.06
	choosing weights	79	4.3	0.49	3.00	5.00	0.06
	fixed rule	43	3.77	1.08	1.00	5.00	0.17
Predictive Validity	holistic	54	0.16	0.19	-0.27	0.42	0.03
	holistic adjustment	93	0.28	0.12	-0.20	0.44	0.01
	choosing weights	79	0.31	0.04	0.19	0.37	0.00
	fixed rule	43	0.35	0.00	0.35	0.35	0.00

The Threat of Technological Unemployment (TOTU)

There was no significant main effect of decision-method on TOTU, ($F(3, 265) = 1.36$, $p = .257$). Neither autonomy-enhanced judgement, nor holistic judgement significantly decrease TOTU compared to pure actuarial judgement. TOTU correlated negatively with use-intentions ($r = -.23$, $p < .01$). Experiencing fear or redundancy was linked to using a particular method less.

Believed Stakeholder Autonomy and Competence Attributions

There was a significant main effect of decision-method on BSAA, (Welch(3, 117) = 3.38, $p = .021$, $\omega^2 = .066$, medium effect). In the fixed-rule condition, BSAA was lower than in the holistic condition ($p = .023$). BSAA was lower in the fixed rule condition than in the choosing weights ($p = .015$) and in the holistic adjustment condition ($p = .021$). The believed stakeholder autonomy attributions, did not differ in the holistic condition compared to the choosing weights ($p = .999$) and holistic adjustment conditions ($p = .989$). Planned contrasts (unequal variances assumed) confirmed that actuarial judgement significantly decreased BSAA compared to holistic judgement $t(62.50) = 2.94$, $p = .005$, $d = .81$ (large effect). The planned contrasts also confirmed that autonomy-enhanced-judgement significantly increased BSAA compared to actuarial judgement, $t(48.57) = 3.17$, $p = .003$, $d = .68$ (medium effect) and not significantly differed from BSAA in the holistic condition, $t(87.60) = -.006$, $p = .995$.

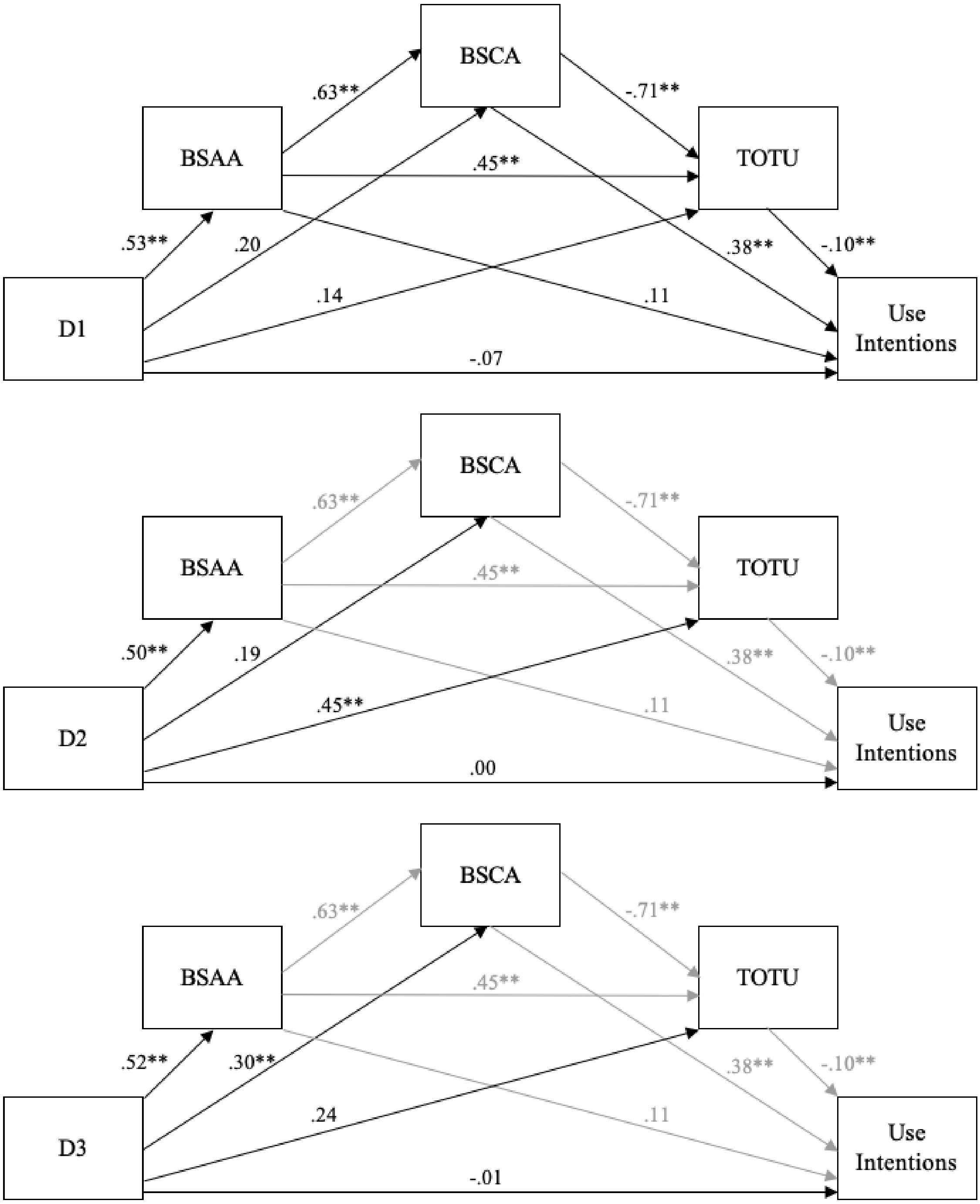
Also, there was a significant main effect of decision-method (holistic, holistic adjustment, choosing weights, fixed rule) on BSCA, (Welch(3, 118) = 3.95, p = .010, ω^2 = .076, medium effect). In the fixed rule condition BSCA were lower than in the holistic condition (p = .006). BSCA were lower in the fixed rule condition than in the choosing weights (p = .019) and in the holistic adjustment condition (p = .030). The believed stakeholder competence attributions, did not differ in the holistic condition compared to the choosing weights (p = .719) and holistic adjustment conditions (p = .604). Planned contrasts (unequal variances assumed) confirmed that actuarial judgement significantly decreased BSCA compared to holistic judgement t(58.42) = 3.54, p =.001, d = .83 (large effect). The planned contrasts also confirmed that autonomy-enhanced-judgement significantly increased BSCA compared to actuarial judgement, t(47.47) = 3.02, p = .004, d = .69 (medium effect) and not significantly differed from BSCA in the holistic condition, t(89.74) = -1.28, p = .204.

Mediation

The mediation analysis revealed that the total direct effects of decision-method on use-intentions were non-significant: D1 (choosing weights vs. fixed condition; b = -.07, t(262) = -.55, p = .585), D2 (holistic adjustment vs. fixed condition; b = .00, t(262) = .03, p = .975) and D3 (holistic vs. fixed condition; b = -.01, t(262) = -.09, p = .931). There were significant results for some relative direct effects (Figure 4). To answer our research, question the relative indirect effects of decsision-method on use-intentions were of interest (Table 5). Particularly the relative specific indirect effect on use-intentions comparing the fixed-rule condition with the choosing weights condition through BSAA, BSCA and TOTU ($D1_{ind7}$= .02, SE = .01, 95% CI[0.00, 0.05]) as well as comparing the holistic adjustment condition with the fixed rule condition ($D2_{ind7}$= .02, SE = .01, 95% CI[0.00, 0.05]). Both were non-significant because the 95% bias corrected CI (lower limit, LLCI; upper limit, ULCI) included zero - the cutoff was two decimals rounded. Even the relative specific indirect effect on use-intentions comparing the fixed-rule condition with the holistic condition through BSAA, BSCA and TOTU was nonsignificant ($D3_{ind7}$= .02, SE = .01, 95% CI[0.00, 0.05]). However, examining relative specific indirect paths exploratory shows that there some significant results; the relative specific indirect effects of decsision-method on use-intentions through BSAA and BSCA was in all comparisons significant ($D1_{ind4}$, $D2_{ind4}$, $D3_{indirect4}$; Table 5). The partially standardized effect sizes for these indirect effects were: $D1_{ind4}$ = .17, SE = .08, 95% CI[0.05, 0.34], $D2_{ind4}$ = .16, SE = .08, 95% CI[0.04, 0.33] and $D3_{ind4}$ = .17, SE = .08, 95% CI[0.04, 0.33]. Lastly, there were also significant results for the relative specific indirect paths of decsision-method on use-intentions through BSCA for the holistic versus fixed rule condition ($D3_{ind2}$; Table 5). The partially standardized effect sizes for this indirect effect was $D3_{indi2}$ = .16, SE = .09, 95% CI[0.03, 0.35].

Figure 4

Relative Direct Effects of the Serial Mediation Analysis



Note. For the Analysis Model Number 6 (Hayes, 2022) was used; 95% C.I.; bootstrap samples 5000; **p< .01 level; * p < .05 level; N=269; Dummy coding: reference group is the fixed-rule condition, D1 compares choosing weights and fixed rule, D2 compares holistic adjustment to fixed rule, D3 compares holistic and fixed rule. Repeated effects are grayed out for better reading purposes.

Table 5

Mediation Analysis: Relative Indirect Effects of Decsision-Method on Use-Intentions

Path	Effect	SE	LLCI	ULCI
1. DM-BSAA-UI				
$D1_{ind1}$: choosing weights vs. fixed	.06	.06	-.04	.18
$D2_{ind1}$: holistic adjustment vs. fixed	.06	.05	-.04	.17
$D3_{ind1}$: holistic vs. fixed	.06	.06	-.04	.18
2. DM-BSCA-UI				
$D1_{ind2}$: choosing weights vs. fixed	.08	.06	-.01	.21
$D2_{ind2}$: holistic adjustment vs. fixed	.07	.06	-.02	.21
$D3_{ind2}$: holistic vs. fixed	.12	.07	.02*	.27*
3. DM-TOTU-UI				
$D1_{ind3}$: choosing weights vs. fixed	-.01	.02	-.06	.03
$D2_{ind3}$: holistic adjustment vs. fixed	-.04	.03	-.11	.00
$D3_{ind3}$: holistic vs. fixed	-.02	.03	-.08	.02
4. DM-BSAA-BSCA-UI				
$D1_{ind4}$: choosing weights vs. fixed	.13	.06	.03*	.26*
$D2_{ind4}$: holistic adjustment vs. fixed	.12	.06	.03*	.25*
$D3_{ind4}$: holistic vs. fixed	.12	.06	.03*	.25*
5. DM-BSAA-TOTU-UI				
$D1_{ind5}$: choosing weights vs. fixed	-.02	.01	-.05	.00
$D2_{ind5}$: holistic adjustment vs. fixed	-.02	.01	-.05	.00
$D3_{ind5}$: holistic vs. fixed	-.02	.01	-.06	.00
6. DM- BSCA-TOTU-UI				
$D1_{ind6}$: choosing weights vs. fixed	.01	.01	.00	.04
$D2_{ind6}$: holistic adjustment vs. fixed	.01	.01	.00	.04
$D3_{ind6}$: holistic vs. fixed	.02	.01	.00	.05
7. DM-BSAA-BSCA-TOTU-UI				
$D1_{ind7}$: choosing weights vs. fixed	.02	.01	.00	.05
$D2_{ind7}$: holistic adjustment vs. fixed	.02	.01	.00	.05
$D3_{ind7}$: holistic vs. fixed	.02	.01	.00	.05

Note. 95% CI, Bootstrap 5000; Labels: DM (decision-method), BSAA (believed stakeholder autonomy attributions), BSCA (believed stakeholder competence attibutions), TOTU (threat of technological unemployment), UI (use-intentions); indirect effects were accepted as statistically significant if the lower limit (LLCI) and upper limit (ULCI) excluded zero.

Discussion

This study aimed to better understand how autonomy in hiring decisions affects people's willingness to use a specific method in practice. Specifically, whether *autonomy-enhanced-judgement* can compete with holistic judgement in terms of use-intention and can increase predictive validity over pure holistic decisions.

Use-Intentions

Our first research question tried to replicate the findings from Nolan and Highhouse (2014) and answer whether autonomy in the decision-method affected use-intentions. The data did not support our hypotheses as the groups did not significantly differ in resistance towards actuarial methods contrary to what has been demonstrated in a vast number of studies (e.g., Eastwood et al., 2012; Highhouse, 2008; Neumann, Niessen et al., 2021; Rynes, 2012) our results did not support the resistance towards actuarial methods (i.e., “algorithm aversion”, Dietvorst, et al., 2015). However, these results are not suggesting that assessment professionals are equally willing to use actuarial judgement. There may be a difference that was not detectable with the current sample, or there could be an alternative explanation. Generally, the effect was in the right direction - use-intentions were highest in the holistic condition and lowest in the fixed rule condition - it just did not reach significance. We would have at least expected to find significant results comparing the holistic and fixed condition. For instance, Nolan and Highhouse (2014) results showed that use-intentions were lowest when both information collection and -combination allowed little autonomy – similar to our fixed-rule condition. Interestingly, Nolan and Highhouse (2014) found that use-intentions were highest the more autonomy the information combination allowed, while little autonomy was permitted in the information collection – like our holistic condition. In our study design, the information was already collected (little freedom in data collection). Since the effect is missing, even for the conditions that should be the furthest apart in use-intentions, another explanation might be plausible for why the groups did not differ.

For transparency purposes, the participants were informed at the beginning of the task that *“although the decision-rule will probably not result in perfect performance predictions, research showed that using such a rule, results in more accurate performance predictions than using one’s intuition and expertise” and “adjusting decision-rule predictions based on one’s intuition and expertise usually decreases prediction accuracy”.* One possible alternative explanation is that the provided information affected use-intentions through 1) educating the participants or 2) creating socially desirable responses. A recent study by Neumann et al. (2021) showed that short educational interventions could increase the use of decision-rules. Furthermore, Eastwood and Luther (2016) found that when participants received information that using a decision-rule improves prediction accuracy over holistic decisions, people reported higher use-intention. Thus, closing the knowledge gap in actuarial judgement might generally receive too little attention given its possible impact (e.g., Banks et al., 2016; Gill, 2018). Research on the topic is sparse (Neumann, Hengeveld et al., 2021), and future studies should investigate how far the *information effect* could affect decision-makers attituded towards actuarial judgement. Also, the provided information might have led participants to assume that they should indicate liking a method more to receive the reward (Dodou & de Winter, 2014).

Another possible explanation is that participants might not have made the transfer from the presented scenario to their actual work-life. This might have resulted in a weak attitude towards using a particular method (Fazio, 1995). Especially because the use-intentions means of the different conditions were all centred around the middle of the scale (Table 4) suggest, participants were irresolute about using any of the methods. Even though we tried to make the scenario relatable and took multiple measures to prevent careless responding (e.g., screeners, response time; Huang & Wang, 2021), we cannot say with certainty that the reported use-intentions would translate to an actual use a in practice. Many studies that found an effect on use-intentions used more interactive lab experiments (e.g., Arkes et al., 1986; Dietvorst et al., 2015) or vividly detailed scenarios (Eastwood et al., 2012), which might have evoked stronger feelings of task identification than an online experiment. Future studies should therefore investigate whether task framing could be a contributing factor and generally be aware that there is a gap in reported intention and actual behaviour (Sheeran, 2002).

Lastly, Nolan (2012) found that even though decision-method (holistic vs actuarial) was related to feelings of competence and autonomy, it was not related to use-intentions. However, use-intentions, were related to the *individual need* for competence and autonomy. This suggests that use-intentions might be also predicted by individual differences in perceived need fulfilment which future studies should keep in mind.

Nonetheless, the use-intentions between the holistic group and the autonomy-enhanced-judgement groups were similar, which also suggest that people would be equally open to using autonomy-enhanced-judgement or holistic judgement. If the predictive validity of autonomy-enhanced-judgement is superior to holistic judgement and use-intentions are equal, it supports autonomy-enhanced-judgement as a promising decision-tool. However, to establish the utility of autonomy-enhanced-judgement further studies need to investigate whether decision-makers have a higher intention to use autonomy-enhanced-judgement compared to actuarial judgement.

Predictive Validity

There is little utility in a method that does not improve predictive validity over pure holistic judgement. Thus, our second research question sought to answer whether autonomy-enhanced-judgement would result in a higher predictive validity than holistic judgement. The results of the present study supported our hypotheses. Both actuarial judgement and autonomy-enhanced-judgement had higher predictive validity than holistic judgement. Self-designing a rule also had a higher predictive validity than holistic adjustment. This pattern of results is consistent with previous findings that holistically adjusting the predictions of a decision-rule decreases predictive validity compared to actuarial judgement (Dawes, 1971; Dietvorst et al., 2018) and literature suggesting that holistic adjustment is more valid than holistic judgement (Dietvorst et al., 2018; Neumann, et al., 2021). Hence, these results further support the idea that people (un-)intentionally anchor their predictions to the predictions of the decision-rule which increases consistency – and consequently prediction accuracy. It would be interesting to investigate whether people are aware of this anchoring effect. Since self-designing a rule is more valid than holistic predictions, it suggests that hiring experts have some understanding about the validity of different predictors and give more weight to relevant information (Yu & Kuncel, 2020). These findings imply utility in using autonomy-enhanced-judgement as an employee selection tool in terms of predictive validity.

The Threat of Technological Unemployment

Nolan & Highhouse (2014) results show that autonomy in the decision process affects use-intentions: acturarial judgement significantly decreased willingness to use a particular method. Furthermore, Nolan et al. (2016) support the relationship between TOTU and use-intentions, showing that fear of redundancy partially explains people's resistance towards actuarial judgement. The current study first investigated how autonomy in the decision-method affects TOTU. However, the results of the present study did not support our hypotheses. Neither autonomy-enhanced judgement nor holistic judgement significantly decreased TOTU compared to pure actuarial judgement. Like the results for use-intentions, the effect might be missing because 1) it might be too small to be detected in this sample or 2) TOTU might not have been evoked by the presented scenario but might occur in a more natural or immersive setting. The fact that perceived autonomy did not correlate with TOTU suggests that the autonomy-restricting nature of actuarial judgement alone is insufficient to elicit a fear of redundancy. However, supporting findings from Nolan et al. (2016), we did find that TOTU correlated negatively with use-intentions suggesting that experiencing TOTU is linked to resistance towards using a particular method.

Believed Stakeholder Attributions

Our fourth research question sought to answer whether believed autonomy and competence attributions from stakeholders differed between autonomy conditions. The results of the present study supported all hypotheses. BSAA and BSCA were lower in actuarial judgement than in holistic judgement and lower in actuarial judgement than in autonomy-enhanced judgement. Furthermore, the data supported our exploratory hypothesis that BSAA and BSCA would be similar in holistic- and autonomy-enhanced judgement. These results suggest that decision-makers are sensitive to how various judgement methods affect stakeholder attributions (supporting Nolan et al., 2016). Decision-makers believed that stakeholders would view them as having less autonomy and competence over the hiring process when actuarial judgement was used to make hiring-decisions compared to holistic judgement. Furthermore, decision-makers believed that stakeholders would view them as having more autonomy and competence over the hiring process when autonomy-enhanced-judgement was used compared to actuarial judgements. This is the first direct demonstration that autonomy-enhanced-judgement can potentially increase decision-makers' beliefs about how autonomous and competent they are perceived by their colleagues and bosses. However, this is the first study to show that decision-makers know how different decision-methods affect stakeholder attributions of competence. According to self-determination theory, people are susceptible to autonomy-, competence-, and relatedness loss (i.e., feeling connected to others) and will take action to regain it (Deci & Ryan, 2000; Radel et al., 2011). The results of our study support that autonomy-enhanced-judgement could pose a promising decision-tool to increase decision-makers beliefs about how autonomous and competent they are perceived by their peers while increasing predictive validity over pure holistic judgement. The current research already involves two of three parts (namely autonomy and competence) of the self-determination theory, which are crucial to understanding and predicting attitudes and behaviour (Deci & Ryan, 2000). It would be interesting to examine how relatedness contributes to this dynamic.

In addition, future research should examine contexts in which stakeholders are more salient during the decision-process. Participants may not have factored in stakeholder reactions because stakeholders might not have been salient during the questions on use intentions. Thus, participants might have thought more about the objectively best method to use rather than other people's opinions when indicting their use-intentions. Since BSAA and BSCA had a significant effect on use-intentions, the effect might be even more prominent in scenarios where stakeholders are more salient. Furthermore, it could be helpful to consider the variety of stakeholders because different stakeholder relationships with the decision-maker (e.g., inferior vs superior colleague) might have differential effects on believed autonomy or competence attributions. BSCA and BSAA might also be altered by factors such as organisational structure (strong vs flat hierarchies) or individual personality traits (e.g., people who are more concerned about other people's beliefs). These factors must be considered before making statements about the utility of autonomy-enhanced judgment in practice.

Mediation

Lastly, it was investigated whether believed stakeholder autonomy and competence perceptions and the threat of technological unemployment are mediators in the relationship between the method of decision-making (autonomy-enhanced judgement, holistic judgement, and no-autonomy judgement) and use-intentions (Figure 1). We were particularly interested in comparing the autonomy-enhanced-judgement conditions with the actuarial judgement condition. However, the mediation hypothesis was not supported. Neither the relative direct effect was significant nor the relative indirect effect of the hypothesized pathway, which applied to all three compared dummy conditions. This means that autonomy-enhanced-judgement did not affect use-intentions through BSAA, BSCA and TOTU compared to actuarial judgement.

However, the relationship was better explained without the threat of technological unemployment. Once TOTU was taken out of the equation (exploratory) the relative indirect effect of decision-method on use-intentions over BSAA and BSCA was significant, also applying to all three compared conditions. This means that the participants beliefs about stakeholders' opinions regarding their autonomy and competence *did* affect the willingness to use a particular method, comparing the holistic condition to the fixed condition. Furthermore, this effect was also found when comparing autonomy-enhanced judgment conditions and the fixed-rule condition. Autonomy-enhanced-judgement did not affect use-intentions over the direct pathway – or did not reach significance at least. However, autonomy-enhanced-judgement improved use-intentions compared to actuarial judgement through the beliefs about what other people involved in the hiring process would think about their autonomy and competence.

These findings suggest that there might be a relationship: enhancing autonomy in the decision-process increased believed stakeholder autonomy perceptions, believed stakeholder competence perceptions and use-intentions. This is the first study investigating this serial relationship; further evidence needs to be gathered to solidify this relationship. This mediation would be interesting to test in a study where the stakes are higher in which decision-makers might be more sensitive to stakeholder attributions compared to a hypothetical scenario. Further research should investigate whether this effect is replicable and can be translated into practice. However, one interpretation of these findings is that people are susceptible to opinions regarding their autonomy and competence of other people involved in the hiring process. Increasing autonomy in the hiring process indirectly affects their willingness to use actuarial judgement. Thus, autonomy-enhanced-judgement could increase use-intentions through the indirect pathway of BSAA and BSCA.

Limitations

There are at least three potential limitations concerning the results of this study. A first limitation concerns the sample; even though we included various screeners, we cannot say with certainty that the people participating in the study were assessment professionals. Mason and Suri (2012) discuss that conducting experiments through MTurk poses similar challenges to any online study, the biggest being that the sample is not representative of the whole (or even the online-) population and thus might impede generalizability. Therefore, we can only make claims about assessment professionals after conducting research with a more controlled sample. Further concerns about conducting experiments online are validity, reliability, and data quality (e.g., Chandler et al. 2014). The generalizability issue applies to various settings and outcomes (Shadish et al. 2002). Therefore, the online experiment is only the first step to assessing the utility of autonomy-enhanced-judgement. Further experimental studies should be conducted in various organisational contexts and industry sectors.

A second potential limitation is that measured predictive validity might not apply to predictive validity found in practice. Psychological distance from a decision (as an online experiment might create) can improve predictive validity. Following construal-level theory, viewing a decision in more abstract- rather than context-specific terms can enhance decision-making because the relevant information can receive greater attention (Fukukura et al., 2013; Trope & Liberman, 2000; Neumann, Hengefelet al., 2021) In real life, hiring decisions are never as abstract as the scenario in the study. This further supports the need for studies conducted in more natural, organizational settings, to investigate the utility of autonomy-enhanced-judgement. However, as this study was first to investigate autonomy-enhanced-judgement as a potential decision-tool it was a relevant first step to shed light on the association between autonomy, stakeholder beliefs and the (potential) willingness to use actuarial decision tools in practice.

Last, there are some general limitations of actuarial decision-making that also apply to autonomy-enhanced-judgement. Specifically, valid predictors need to be used that are measured on the same scale (Dawes & Corrigan, 1974). This might sometimes further limit the already meagre enthusiasm for actuarial judgement when assessment professionals are required to exert more effort than for holistic judgement. However, as the extra effort required in using autonomy-enhanced-judgement is outweighed by its potential benefits in predictive validity (which need to be further supported) researcher should continue to popularize any decision-method that might improve hiring decisions.

Findings and Implications

The present study had three key findings. First, autonomy-enhanced-judgement increased predictive validity over holistic judgement. Second, people seemed aware of how autonomous and competent stakeholders would perceive them, and autonomy-enhanced-judgement improved these perceptions. Third, the study revealed a potential serial mediation effect of believed stakeholder attributions of autonomy and competence in the relationship between decision-autonomy and use-intentions. Autonomy-enhanced-judgement improved use-intentions through the beliefs about other people's beliefs involved in the hiring process compared to actuarial judgement. So far, this study supported that assessment professionals equally accept autonomy-enhanced-judgement as holistic judgement, but further evidence needs to be gathered to reach a definite conclusion.

This research can be seen as a first step towards answering how we can make personnel selection methods more valid and fair. Striving for fairness and evidence-based practices is particularly relevant in personnel selection to promote a society where people are ensured equal opportunities (Castilla, 2016). However, holistic decisions do not provide the necessary transparency to discover whether a decision is indeed based on an applicant's merit rather than irrelevant personal characteristics or biases. Holistic judgement can often provide an arcane method to sustain the organisations' culture and perpetuate implicit and explicit beliefs of the people who work there (Simplicio, 2007). Because holistic judgement is most widely used, hiring decisions are seldom based on true meritocracy (Castilla & Benard, 2010; Simplicio, 2007). Instead, current hiring practices favour some candidates while being biased against others (Petersen & Togstad, 2006; Tosi & Einbender, 1985; Zschirnt & Ruedin, 2016). Thus, in addition to improving the validity of decisions, actuarial judgement has another advantage: transparency. Actuarial judgement exposes the decision process allowing it to be evaluated and altered and can make errors and biases evident. Implementing decision-methods that approximate actuarial judgement might thus increase fairness in hiring decisions. Even though holistic-adjustment has similar issues with transparency, designing decision-rules is a promising method to increase predictive validity and transparency. Thus, both autonomy-enhanced-judgement methods show promising advantages over the current status quo in making hiring decisions.

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

Arthur Jr, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*(2), 435-442. https://doi.org/10.1037/0021-9010.93.2.435

Banks, G. C., Pollack, J. M., Bochantin, J. E., Kirkman, B. L., Whelpley, C. E., & O'Boyle, E. H. (2016). Management's science–practice gap: A grand challenge for all stakeholders. *Academy of Management Journal, 59*(6), 2205–2231. https://doi.org/10.5465/amj.2015.0728

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217. https://doi.org/10.1037/h0047470

Callinan, M. & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*, 248-260. https://doi.org/10.1111/1468-2389.0015

Castilla, E. (2016). *Achieving meritocracy in the workplace*. MIT Sloan Management Review. 35–41. http://mitsmr.com/1PZ7SPl

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods*, *46*(1), 112-130. https://doi.org/10.3758/s13428-013-0365-7

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press. 267–401 https://doi.org/10.4324/9780203771587

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, 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x

Davis, M. H., Soderlund, T., Cole, J., Gadol, E., Kute, M., Myers, M., & Weihing, J. (2004). Cognitions associated with attempts to empathize: How do we imagine the perspective of another?. *Personality and Social Psychology Bulletin*, *30*(12), 1625-1635. https://doi.org/10.1177/0146167204271183

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., & Corrigan, B. (1974). Linear models in decision making. *Psychological bulletin*, *81*(2), 95-106. https://doi.org/10.1037/h0037613

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

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134. https://doi.org/10.1016/0092-6566(85)90023-6

Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media. 35-42.

Diab, D.L., Pui, S., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in U.S. and Non-U.S. Samples. *International Journal of Selection and Assessment, 19*(2), 209-216. https://doi.org/10.1111/j.1468-2389.2011.00548.x

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114-126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155-1170. https://doi.org/10.1287/mnsc.2016.2643

Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36*, 487-495. https://doi.org/10.1016/j.chb.2014.04.005

Eastwood, J., & Luther, K. (2016). What you should want from your professional: The impact of educational information on people's attitudes toward simple actuarial tools. *Professional Psychology: Research and Practice*, *47*(6), 402-412. https://doi.org/10.1037/pro0000111

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458-468. https://doi.org/10.1002/bdm.741

Farr, J. L., & Tippins, N. T. (Eds.). (2017). *Handbook of employee selection*. Taylor & Francis.365-397. https://lccn.loc.gov/2016039053

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. *Attitude strength: Antecedents and consequences*, *4*, 247-282.

Fukukura, J., Ferguson, M. J., & Fujita, K. (2013). Psychological distance can improve decision making under information overload via gist mem- ory. Journal of Experimental Psychology: General, 142(3), 658–665. https://doi.org/10.1037/a0030730

Gill, C. (2018). Don't know, don't care: An exploration of evidence-based knowledge and practice in human resource management. Human Resource Management Review, 28(2), 103–115. https://doi.org/10.1016/j.hrmr .2017.06.001

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, *73*(6), 422-432. https://doi.org/10.1037/h0029230

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*(2), 293-323. http://dx.doi.org.proxy-ub.rug.nl/10.1037/1076-8971.2.2.293

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A* *regression-based approach* (3rd edition). New York: The Guilford Press. 77-229. https://lccn.loc.gov/2021031108

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

Himmelfarb, S., & Anderson, N.H. (1975). Integration-theory applied to opinion attribution. Journal of Personality and Social Psychology, 31, 1064-1071. https://doi.org/10.1037/h0076955

Huang, J. L., & Wang, Z. (2021). Careless Responding and Insufficient Effort Responding. In Oxford Research Encyclopedia of Business and Management. https://doi.org/10.1093/acrefore/9780190224851.013.303

Karelaia, N., & Hogarth, R. (2008). Determinants of linear judgement: a meta-analysis of lens model studies. *Psychol. Bull.*134, 404–426. https://psycnet.apa.org/doi/10.1037/0033-2909.134.3.404

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

Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, *28*(2), 107-128. https://doi.org/10.1037/h0034225

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

Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt, N. Schmitt (Eds.), *The Oxford handbook of personnel assessment and selection* (pp. 383-410). New York, NY, US: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0017

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods, 44(*1), 1-23. https://doi.org/10.3758/s13428-011-0124-6

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50*(3), 370–375. https://doi.org/10.1207/s15327752jpa5003_6

Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A tutorial on mechanical decision-making for personnel and educational selection. *Frontiers in psychology*, *10*, 1-8. https://doi.org/10.3389/fpsyg.2019.03002

Neumann, M., Hengeveld, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2021). Education increases decision-rule use: An investigation of education and incentives to improve decision making. Advance online publication. *Journal of Experimental Psychology: Applied*, *28*(1), 166-178. https://doi.org/10.1037/xap0000372

Neumann, M., Niessen, A. S. M., & Meijer, R. R. (2021). Implementing evidence-based assessment and selection in organizations: A review and an agenda for future research. *Organizational Psychology Review*, *11*(3), 205-239. https://doi.org/10.1177/2041386620983419

Neumann, M., Niessen, A. S. M., Tendeiro, J. N., & Meijer, R. R. (2022). The autonomy-validity dilemma in mechanical prediction procedures: The quest for a compromise. *Journal of Behavioral Decision Making*, *35*(4), e2270. https://doi.org/10.1002/bdm.2270

Nolan, K. P. (2012). *Basic psychological need fulfilment and user resistance to objective and analytical decision-making practices in employee selection (Doctoral dissertation, Bowling Green State University).* http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1343479006

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*(1), 30-47. https://doi.org/10.25035/pad.2016.004

Nolan, K. P., Dalal, D. K., & Carter, N. (2020). Threat of technological unemployment, use intentions, and the promotion of structured interviews in personnel selection. *Personnel Assessment and Decisions, 6*, 38–53. https://doi.org/10. 25035/pad.2020.02.006

Nolan, K. P., & Highhouse, S. (2014). Need for autonomy and resistance to standardized employee selection practices. *Human Performance*, *27*(4), 328-346. https://doi.org/10.1080/08959285.2014.929691

Petersen, T., & Togstad, T. (2006). Getting the offer: Sex discrimination in hiring. *Research in Social Stratification and Mobility*, *24*(3), 239-257. https://doi.org/10.1016/j.rssm.2006.06.001

Radel, R., Pelletier, L. G., Sarrazin, P., & Milyavskaya, M. (2011). Restoration process of the need for autonomy: the early alarm stage. *Journal of Personality and Social Psychology*, *101*(5), 919-934. https://doi.org/10.1037/a0025196

Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and social Psychology*, *42*(6), 1137-1145. https://doi.org/10.1037/0022-3514.42.6.1137

Rynes, S. L. (2012). The research-practice gap in I/O psychology and related fields: Challenges and potential solutions. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (Vol. 1., pp. 409–452). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199928309.001.0001

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*(3), 178–200. https://doi.org/10.1037/h0023624

Schlegel, D., Kolb, S. J., Luciano, J. M., Tovar, J. M., Cucchiara, B. L., Liebeskind, D. S., & Kasner, S. E. (2003). Utility of the NIH Stroke Scale as a predictor of hospital disposition. *Stroke*, *34*(1), 134-137. https://doi.org/10.1161/01.str.0000048217.44714.02

Sekiguchi, T. (2004). Person-organization fit and person-job fit in employee selection: A review of the literature. *Osaka keidai ronshu*, *54*(6), 179-196.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs* *for generalized causal inference.* Houghton, Mifflin and Company. 5-27. https://www.alnap.org/system/files/content/resource/files/main/147.pdf

Sheeran, P. (2002). Intention—behavior relations: a conceptual and empirical review. European review of social psychology, 12(1), 1-36. https://doi.org/10.1080/14792772143000003

Simola, S. K., Taggar, S., & Smith, G. W. (2007). The employment selection interview: disparity among research-based recommendations, current practices and what matters to Human Rights Tribunals. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *24*(1), 30-44. https://doi.org/10.1002/cjas.4

Simplicio, J. S. (2007). A closer look at the truth behind the hiring process: How colleges really hire. *Education*, *128*(2), 256-262.

Schmitt, N., Chan, D., & Chan, E. (1998). *Personnel selection: A theoretical approach*. Sage. 5-12.

Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *Academy of Management Journal*, *28*(3), 712-723. https://doi.org/10.5465/256127

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*(6), 876–889. https://doi.org/10.1037/0022-3514.79.6.876

Yu, M. C., & Kuncel, N. R. (2020). Pushing the limits for judgmental consistency: Comparing random weighting schemes with expert judgments. *Personnel Assessment and Decisions*, *6*(2), 1-10. https://doi.org/10.25035/pad.2020.02.002

Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests. 1990–2015. *Journal of Ethnic and Migration Studies*, *42*(7), 1115-1134. https://doi.org/10.1080/1369183x.2015.1133279

Appendix A: R Script for the Applicant Selection from the Stimulus Dataset

```
#####################################################################################
# Improving Stakeholder Perceptions in Mechanical Judgment Procedures - Study 2 #

# Author: Marvin Neumann

# Content of the script: Selecting 40 applicants from the Kausel dataset that contains
# 236 applicants to retain the correlations as much as possible
#####################################################################################

# set working directory -------------------------------------------------------------------
# setwd("~/Documents/Groningen/PhD/Teaching/Master thesis supervision/Isabell Budzynski")

# install and load packages ---------------------------------------------------------------
# install.packages("tidyverse")
# install.packages("psych")
# install.packages("scales")

library(tidyverse) # load a package that makes many things much easier
library(psych)     # needed to obtain regression weights from the meta-analysis
library(scales)    # needed for rescaling

# loading Kausel applicant data -----------------------------------------------------------
# make sure the following CSV file is stored in your working directory
data <- read.csv2("base_overconfid_study.csv", dec = ",")

# recoding and renaming data --------------------------------------------------------------
data$Gmean <- as.numeric(data$Gmean)

### Variable explanation ###
# ID_Number   = Person ID
# gen_perf    = job performance rating (criterion)
# Gmean       = Score on a cognitive ability test (% of answers right)
```

```
# Consc       = Average score on a conscientiousness test
# Interview   = Score on an interview (theoretical range: 1-5, actual range: 2-5)

# Kausel data renamed
data <- data %>%
   rename(ID = ID_Number,
          Criterion = gen_perf,
          GMA = Gmean,
          Consc = CONSC,
          Interview = interview) %>%
   select(ID,GMA,Consc,Interview,Criterion)

# correlations among full applicant dataset -----------------------------------------------
cormat <- cor(data[,-1])

# Applicant-selection algorithm ----------------------------------------------------------

# we create an empty correlation matrix with the same dimensions as cormat
dev <- matrix(data = 0, nrow = 4, ncol = 4)

# On all offdiagonals, we define a maximum deviation we would tolerate.
# Here I say that in the 40 applicants we will select, I want the correlations between
# variables to not differ by more than 0.015 compared to the full dataset (236 applicants)
dev[lower.tri(dev)] <- dev[upper.tri(dev)] <- 0.015

# Now we create an algorithm that samples cases from the full dataset and then evaluates
# whether the deviation between the correlations based on the full dataset
# and the correlations in the 40 applicants is less or equal to dev = 0.015.
# If TRUE, the algorithm stops. If FALSE, a new set of 40 applicants is sampled.
# I tried this several times and the smallest margin for which the algorithm finds a solution
# seems to be 0.015. That's quite good. So, in our applicant dataset, no correlation is
# different by more than 0.015 points compared to the full applicant dataset.

dummy <- TRUE
while (dummy) {
   idx <- sample(x = 1:nrow(data), size = 40)
   samp <- data[idx, ]
```

```
    samp_cormat <- cor(samp[,-1])
    dummy <- ifelse(all(abs(cormat - samp_cormat) <= dev), FALSE, TRUE)}
# To see that it worked, you can compare cormat (correlations based on all 236 cases)
# and samp_cormat (correlations based on our 40 selected applicants)

# Our final 40 applicants are saved in the data.frame "samp"!

# Adding decision-rule predictions based on meta-analytic weights ----------------------------

# Recreating the matrix from the meta-analysis by Cortina et al. (2000)
cortina.mat <- matrix(data = c(1, .07, .06, .45,
                               .07, 1, .09, .27,
                               .06, .09, 1, .20,
                               .45, .27, .20, 1),
                      nrow = 4,
                      ncol = 4)
rownames(cortina.mat) <- colnames(cortina.mat) <- c("GMA","Consc","Interview","Criterion")

# obtaining standardized regression weights with the setCor function from the psych package
weights <- setCor(Criterion ~ GMA + Consc + Interview, data = cortina.mat)$coefficients

# So, standardized regression weights are (rounded) GMA = 0.42, Consc = 0.23, Interview = 0.15

# Standardize applicant data ---------------------------------------------------------------
# Now we standardize our predictors from the data from our 40 applicants we selected

std.data <- data.frame(
    ID = samp$ID,            # we don't standardize the ID variable, which would make no sense
    std.GMA = scale(samp$GMA),
    std.Consc = scale(samp$Consc),
    std.Interview = scale(samp$Interview)
)
# Now we multiply the standardized predictors by the standardized regression weights
# I call this standardized composite score
std.comp <- as.matrix(std.data[2:4])%*%as.numeric(weights)

# rescaling decision-rule predictions ------------------------------------------------------
```

```
# Now we rescale this standardized composite score to our 5 point scale
# I call this decision-rule prediction
# Furthermore, we round this rule prediction to one decimal.
rule.preds <- rescale(std.comp,
                   to = c(1, 5)) %>% round(1)

# Create data frame for Qualtrics ------------------------------------------------------------
# Lastly, we merge all our data of interest into one data.frame that we can work with in Qualtrics
Qualtrics.data <- data.frame(
  samp,
  std.data[,2:4],
  rule.preds
)
# As a check, we can correlate our rule predictions with the criterion, and we see r = 0.35471.
# So this worked as intended
cor(Qualtrics.data$rule.preds, Qualtrics.data$Criterion)

# Write CSV ---------------------------------------------------------------------------------
# Finally, save our data frame as a CSV file
write.csv(Qualtrics.data, file = "Qualtrics.applicant.data.csv")
```

+

Appendix B: Screeners

Screener 1

Please indicate up to five tasks in which you most frequently engage at work.

- **Staffing organizational units -** Recruiting, interviewing, selecting, hiring, and promoting applicants/employees.
- **Selling or influencing others -** Convincing others to buy merchandise/goods.
- **Guiding, directing, and motivating subordinates -** Providing guidance and direction to subordinates, including setting performance standards and monitoring performance.
- **Controlling machines and processes -** Using either control mechanisms or direct physical activity to operate machines or processes (not including computers or vehicles).
- **Documenting/recording information -** Entering, transcribing, recording, storing, or maintaining information in written or electronic form.
- **Repairing and maintaining equipment -** Servicing, repairing, adjusting, and testing machines, devices, moving parts, and equipment that operate on the basis of mechanical or electrical principles.
- **Scheduling work and activities -** Scheduling events, programs, and activities, as well as the work of others.
- **Making decisions and solving problems** - Analyzing information and evaluating results to choose the best solution and solve problems.
- **Judging the qualities of things, services, or people -** Assessing the value, importance, or quality of things or people.
- **Estimating the quantifiable characteristics of products, events, or information -** Estimating sizes, distances, and quantities; or determining time, costs, resources, or materials needed to perform a work activity.
- **Getting information -** Observing, receiving, and otherwise obtaining information from all relevant sources.
- **Monitoring processes, materials, or surroundings -** Monitoring and reviewing information from materials, events, or the environment, to detect or assess problems.

Screener 2

Approximately, how many hiring decisions do you undertake in a year?

- 0
- 1 - 5
- 6 – 10
- 11 – 20
- more than 20

(If 0 was selected the study was discontinued)

Appendix C: Task

General Task Information and Bonus Scheme

In the following task, we will ask you to make 40 performance predictions **based on real-life data.** An airline was opening new offices and filled vacancies for the job of a ticket agent. As part of the selection procedure, applicants took a **general mental ability test** and filled in a standardized **personality questionnaire (conscientiousness questionnaire)**. They were also interviewed for the job via an **unstructured interview conducted by a line manager.** All applicants were hired, except for those who obtained the lowest interview rating possible. Three months after being hired, the applicants were assessed by their supervisors on their overall job performance. Thus, we already know how the applicants performed later on the job. In this task, we ask you to predict applicants' later job performance.

The information will be presented to you on the following scales:

1. The scores from the **general mental ability test** are shown on a scale from 0-100. This score reflects the percentage of correct answers on the test.
2. The **conscientiousness questionnaire** scores are shown on a scale from 1-5, with 5 being the highest conscientiousness score.
3. The scores from the **unstructured interview** range from 2 to 5, with 2 being the lowest and 5 being the highest interview score. The 1 is absent because, per company policy, those scoring 1 were not hired.

With your expertise in making hiring decisions, we now would like you to predict how the applicants will perform on the job. In the following task, we ask you to make **40 performance predictions** based on real data. We would like to know how well, you as a hiring expert, can estimate how applicants will perform on the job. You will receive some information about the applicants, and for each applicant, **you predict how that applicant will perform on the job assessment three months later.** Remember, for performing well on this task, you will be able to earn an additional **monetary reward (up to 5$)** for accurate performance predictions. The table below shows the relation between prediction accuracy and reward. For example, if your 40 performance predictions deviate on average between 0.00 and 0.19 from the applicants' actual job performance, you will earn 5$. If your predictions are off by 0.4 to 0.59 on average, you will earn 3$.

Reward	Average deviation from applicants' actual job performance
5 $	less than 0.2
4 $	0.2 - 0.39

Reward	Average deviation from applicants' actual job performance
3 $	0.4 - 0.59
2 $	0.6 - 0.79
1 $	0.8 - 1.0

We would like you to make the performance predictions by using a specific method to combine the information from the applicants.

Task Information: Holistic Condition

Please review the scores of the applicants and predict **based on your intuition and expertise** how the applicants will perform on the job.

Task Information: Holistic Adjustment

In this approach, we show you, for each applicant, the performance prediction of a decision rule. An assessment professional designed this decision rule based on numerous empirical research findings. The decision rule looks like this:

Decision-rule prediction = General mental ability test score * **53** + Conscientiousness questionnaire score * **28** + Interview rating * **19**.

As you can see above, the decision rule assigns the following weights to the information:

- General Mental Ability Test: **53**
- Conscientiousness Questionnaire: **28**
- Interview Rating: **19**

The chosen weights correspond to the importance assigned to each piece of the information. So, based on the numerous empirical research findings, the assessment professional decided to give most weight to the general mental ability test score, and least weight to the interview rating. Above you can see that the scores and ratings of an applicant were multiplied by weights (*) and then added up (+).

The higher the decision-rule prediction, the more likely it is that the applicant shows good job performance.

You can use the decision-rule prediction in different ways. You can review the scores of the applicants and consult the prediction of the decision rule shown below and decide **based on your intuition and expertise** how you would like to combine the information to predict the applicants' job performance. If you only want to use the exact decision-rule prediction, you simply reproduce this prediction as your answer.

Although the decision rule will probably not result in perfect performance predictions, research showed that adjusting decision-rule predictions based on one's intuition and expertise usually decreases prediction accuracy.

Please review the scores of the applicants and predict **based on your intuition and expertise** how the applicants will perform on the job. You may consult the **prediction of the decision rule.**

Task Information: Choosing Weights Condition

In this approach, we want you to **design a decision rule.** You will estimate how important each piece of information is for later job success and create **one rule** that will be used to make performance predictions **for all applicants.** The higher the decision-rule prediction, the more likely it is that the applicant shows good job performance.

Before you get descriptions of the applicants, you will decide how much weight you will assign to each piece of information: **General Mental Ability Test, Conscientiousness Questionnaire, and Interview Rating.**

The chosen weights correspond to the importance you assign to each piece of information. To illustrate this, if you think interview ratings are most important in predicting job performance you should weigh it more heavily than the other information. Conversely, if you think the general mental ability test or conscientiousness questionnaire will be most important in predicting later job success you should assign more weight to them respectively.

On the basis of your designed decision rule an overall score for each applicant will be calculated. Below you can see that the scores and ratings of an applicant are multiplied by weights (*) and then added up (+). The higher the overall score of the applicant the higher the chance of job success.

The decision rule looks like this:

Decision-rule prediction = General mental ability test score * your chosen weight 1 + Conscientiousness questionnaire score * your chosen weight 2 + Interview rating * your chosen weight 3

You will see the predictions made based on your decision rule. But once you designed the rule you cannot adjust its predictions.

Although your decision rule will probably not result in perfect performance predictions, research shows that using such a rule results in more accurate performance predictions than using one's intuition and expertise.

First, you will design **one decision rule** that will make performance predictions for **all applicants**. Your decision rule will then predict the job performance of the applicants. To demonstrate how your rule operates, we will show you 10 performance predictions.

Task Information: Fixed-Rule Condition

In this approach, we show you, for each applicant, the performance prediction of a decision rule. An assessment professional designed this decision rule based on numerous empirical research findings. The decision rule looks like this:

Decision-rule prediction = General mental ability test score * **53** + Conscientiousness questionnaire score * **28** + Interview rating * **19**.

As you can see above, the decision rule assigns the following weights to the pieces of information:

- General Mental Ability Test: **53**
- Conscientiousness Questionnaire: **28**
- Interview Rating: **19**

The chosen weights correspond to the importance assigned to each piece of information. So, based on the numerous empirical research findings, the assessment professional decided to give most weight to the general mental ability test score, and least weight to the interview rating. Above you can see that the scores and ratings of an applicant were multiplied by weights (*) and then added up (+).

The higher the decision-rule prediction, the more likely it is that the applicant shows good job performance.

You will see the predictions made by the decision rule. The applicants' predicted job performance will be shown to you, and you **cannot adjust this prediction** based on your intuition and expertise. The decision rule will predict the performance of all 40 applicants. To demonstrate how the rule operates, we will show you 10 performance predictions which you cannot adjust.

Although the decision rule will probably not result in perfect performance predictions, research showed that adjusting decision-rule predictions based on one's intuition and expertise usually decreases prediction accuracy.

Appendix D: Measures

Questions were answered on a 5-point Likert scale: Strongly disagree (1), Somewhat disagree (2), Neither agree nor disagree (3), Somewhat agree (4), Strongly agree (5). (R) stands for reverse coded items.

General Instructions

As mentioned earlier, an airline was filling vacancies for the job of a ticket agent. Please imagine that **the airline used the approach you just used** to make performance predictions. **The applicants with the highest performance prediction scores were hired.**

Please also imagine that **you were the manager in charge** to deliver the decision to the applicants.

Please indicate in the following questions the extent to which you agree or disagree with a statement

Use Intentions

- If I were in charge, I would use this approach to make hiring decisions.
- If I could use a different approach to make hiring decisions, I would. **(R)**
- I would choose to use this approach to make future hiring decisions.

TOTU

Consistently using this approach to make hiring decisions would …

- Undermine my usefulness as an employee.
- Reduce the perceived importance of my position within the company.
- Lessen others' beliefs about the value I provide to my employing organization.
- Diminish my professional reputation.
- Threaten the status of my employment within my organization.

Perceived Autonomy

Using this approach to make hiring decisions would give me a sense of ...

- Control
- Choice
- Free Will
- Influence
- Self-sufficiency
- Freedom

Locus of Causality

Please think about the other people at your organisation who are familiar with the process used to make this hiring decision…

- consider me responsible for the outcome of the decision.

- think the outcome of this hiring decision reflects on my ability to make hiring decisions.
- attribute the outcome of this hiring decision to me.

Personal Control

Think about the other people at your organisation who are familiar with the process used to make this hiring decision. **The others** would ...

- think I had control over how the hiring decision was made.
- think I had the power to decide which candidate was hired.
- think I was able to change the hiring process as I saw fit.

Believed Stakeholder Competence Attribution

Think about the other people at your organisation who are familiar with the process used to make this hiring decision. How would they appraise your competence? **The others** would think I am...

- effective
- capable
- useful
- skillful
- competent
- accomplished