Analyzing anchoring bias in attribute weight elicitation of SMART, Swing, and best-worst method

Rezaei, Jafar; Arab, Alireza; Mehregan, Mohammadreza

DOI
10.1111/itor.13171

Publication date
2022

Document Version
Final published version

Published in
International Transactions in Operational Research

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 0 (2022) 1–31
DOI: 10.1111/itor.13171

Analyzing anchoring bias in attribute weight elicitation of SMART, Swing, and best-worst method

Jafar Rezaei[a,*], Alireza Arab[b] and Mohammadreza Mehregan[b]

[a] *Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands*
[b] *Faculty of Management, University of Tehran, Tehran, Iran*
E-mail: j.rezaei@tudelft.nl [Rezaei]; alireza.arab@ut.ac.ir [Arab]; mehregan@ut.ac.ir [Mehregan]

Received 2 July 2021; received in revised form 27 May 2022; accepted 20 June 2022

Abstract

In this study, the existence of anchoring bias—people's tendency to rely on, evaluate, and decide based on the first piece of information they receive—is examined in two multi-attribute decision-making (MADM) methods, simple multi-attribute rating technique (SMART), and Swing. Data were collected from university students for a transportation mode selection. Data analysis revealed that the two methods, which have different starting points, display different degrees of anchoring bias. Statistical analyses of the weights obtained from the two methods show that, compared to Swing (with a high anchor), SMART (with a low anchor) produces lower weights for the least important attributes, while for the most important attributes, the opposite is true. Despite their differences in anchoring bias, analytical approaches supported by empirical studies suggest that both methods (SMART and Swing) overweigh the less important attributes and underweigh the more important attributes. As such, we examined whether the best-worst method (BWM), which has two opposite anchors in its procedure (a possible promising anchoring debiasing strategy), could produce results that are less prone to anchoring bias. Our findings show that the BWM is indeed able to produce lower weights (compared to SMART and Swing) for the less important attributes and higher weights for the more important attributes. This study shows the vulnerability of MADM methods with a single anchor and supports the idea that MADM methods with multiple (opposite) anchors, like BWM, are less prone to anchoring bias.

Keywords: cognitive bias; anchoring bias; multi-attribute weighting; MADM; SMART; Swing; best-worst method

1. Introduction

Multi-attribute decision-making (MADM), also called multi-criteria decision-making (MCDM), involves evaluating different alternatives (options) with respect to certain attributes (criteria) with the ultimate aim of ranking, sorting, or selecting the alternatives. In any MADM problem, a list

*Corresponding author.

of alternatives and a list of attributes must be identified. The attributes not only improve decision-makers' (DMs) ability to define the alternatives, but their relative importance also plays a crucial role in formulating and solving the problems. The main subject of this study is the identification of the relative importance (weight) of the attributes. Various MADM weighting methods have been developed in recent decades, most of which are based on DMs' evaluation and subjective judgment concerning the weights of the attributes. As Tversky and Kahneman (1974) explained, people's subjective judgments rely on a limited number of heuristic principles that reduce the complex tasks involved. Generally, heuristics help simplify the decision-making process. However, in some cases, they lead to error and deviations from rational decision-making, as a result of which the optimal results in a problem are distorted (Bazerman and Moore, 1994). The errors involved are known as cognitive biases, and to date, researchers have identified many cognitive biases, and the subjective process of eliciting weights in MADM problems is prone to these cognitive biases. We need to examine these biases in MADM weighting methods and provide effective debiasing solutions to reduce their impact on the final results.

Despite the rich literature on cognitive biases in behavioral psychology, few researchers have so far discussed cognitive biases in the MADM field (Montibeller and von Winterfeldt, 2015b; Marttunen et al., 2018; Rezaei, 2021). Range insensitivity bias (Gabrielli and von Winterfeldt, 1978; Von Nitzsch and Weber, 1993; Fischer, 1995; Pöyhönen and Hämäläinen, 2000; Lin, 2013), proxy bias (Fischer et al., 1987), equalizing bias (Rezaei et al., 2022), splitting bias (Borcherding and von Winterfeldt, 1988; Weber et al., 1988; Pöyhönen and Hämäläinen, 1998, 2000, 2001; Jacobi and Hobbs, 2007; Hämäläinen and Alaja, 2008), framing bias, loss aversion, and status quo (Deniz, 2020), and anchoring bias (Corner and Buchanan, 1997; Rezaei, 2021) are among the cognitive biases that researchers have examined within the context of MADM. There are also some comprehensive studies on the theoretical implications of cognitive biases in MADM, which highlight the need to conduct empirical and experimental analyses and examine different biases in real-world decision-making problems (Montibeller and von Winterfeldt, 2015a, 2015b, 2018; Montibeller, 2018).

Of the cognitive biases, anchoring bias has been shown to be a more visible and crucial bias in MADM weighting methods. Anchoring bias, also called "anchoring and adjustment," is a bias that occurs when people make estimates by starting from an initial value and adjusting (which is usually insufficient) that value to reach the final answer, and the results are biased toward the initial values (Tversky and Kahneman, 1974). All MADM weighting methods have their own execution procedure for DM evaluation, which leads to different starting points and the associated anchoring bias. In this research, we examine the cognitive bias in MADM weighting methods by looking at a real decision-making problem and providing a debiasing solution to mitigate anchoring bias. To that end, we examined the problem using two well-known MADM weighting methods, the simple multi-attribute rating technique (SMART) (Edwards, 1977) and Swing (von Winterfeldt and Edwards, 1986). The reason we selected these particular methods is that they have opposite starting points (anchors), making them interesting subjects for investigating anchoring bias. SMART has a lower bound anchor and starts with the least important attribute, while Swing has an upper bound anchor and starts with the most important attribute. We propose a hypothesis to test the opposite behavior of these two methods with respect to anchoring bias and test that hypothesis using experimental analysis, after which, following previous studies that have suggested that methods that involve "multiple and counter-anchors" or "consider-the-opposite-strategy" could remedy anchoring bias (Montibeller and von Winterfeldt, 2015a, 2015b, 2018; Rezaei, 2021), we chose best-worst method (BWM), a method that meets that important qualification, as a remedy for anchoring bias and formulated another hypothesis, which we also tested using the experimental analysis, to examine its effectiveness. Rezaei (2021) argued that the two-vector mechanism that exists in the BWM might help cancel out the impact of anchoring bias and proposed examining the effectiveness of its inherent anchoring debiasing feature in an experimental study. Recently, he has shown how the two-vector mechanism actually cancels out the effect of anchoring bias, which occurs in every single vector (Rezaei, 2022).

The main contribution of this study is twofold. First, we investigate anchoring bias in SMART and Swing in a real-world decision-making problem and find that the two methods have different directions of anchoring bias. Second, we examine the debiasing power of the BWM, which could be a remedy to the anchoring bias found in single anchor methods. We found that the BWM, which has two (opposite) anchors, could lead to less biased conclusions. Finally, this study provides significant insights into the mechanism of anchoring bias in MADM methods.

In Section 2, the theoretical background of anchoring bias, its main causes, debiasing strategies, and anchoring bias in MADM are discussed. In Section 3, the MADM weighting methods used in this research are described. In Section 4, the research hypotheses are formulated, and the experiment design is discussed. Data analysis and discussion are provided in Section 5, and finally, Section 6 presents the conclusion and future research suggestions.

2. Theoretical background

In this section, we start by discussing anchoring bias, after which we look at the main causes of anchoring bias and the main debiasing strategies from the behavioral psychology literature. Finally, we discuss studies that have addressed anchoring bias within the context of MADM.

2.1. Anchoring bias

The word "cognitive" is derived from the Latin word "cognate," meaning consciousness, which is rooted in the concept of bounded rationality proposed by Simon (1957). DMs usually have limited time, money, and information at their disposal when faced with a decision-making problem, which is why they look for a satisfactory solution rather than an optimal and fully rational one. Tversky and Kahneman continued Simon's work, providing details of the systematic biases that affect people's judgments, and their efforts have led to our understanding of the judgments and cognitive biases that play a role in decision-making (Tversky and Kahneman, 1974). They found that people rely on simplifying strategies and rules of thumb in their decision-making and called these strategies "heuristics." Although these rules reduce the complex tasks of the decision-making process due to time, cost, information, and the DM's cognitive ability limitations, in some cases, they lead to cognitive biases and ultimately sub-optimal decisions. So far, many cognitive biases have been identified, with anchoring bias being one of the most important, widely studied, and well-known cognitive biases (Ünveren and Baycar, 2019), one that was first discussed by Tversky and Kahneman (1974). Anchoring bias refers to the fact that DMs will place more importance on an initial value in their judgment, after which they will try to adjust the initial value to arrive at a more meaningful evaluation. However, the adjustments are usually insufficient, which is why this bias is also called "anchoring and adjustment." Researchers have identified two ways in which anchors (initial values) affect the anchoring and adjustment process, named "with explicit direction" and "without explicit direction." In the first way, people's attention is explicitly directed toward anchors by the information that is provided first. For example, Tversky and Kahneman (1974) conducted an experiment and asked subjects to estimate the proportion of United Nations member states in African nations. Subjects were randomly assigned to 10% or 65% by spinning a wheel of fortune, with numbers ranging from 0% to 100%. Then, they asked people to name the actual percentage. The results showed that the randomly assigned percentages affected people's responses. Subjects who had been given a low anchor (10%) gave a lower response (25% on average) than those who were given a high anchor (65%), who gave a higher response (45% on average). However, the second way anchors can occur involves incidental, informative, or self-generated anchors. For example, Critcher and Gilovich (2008) asked subjects to estimate the percentage of new phone sales with the model numbers "P17" and "P97." The results showed that people's sales forecasts were affected by the incidental anchor contained in the model number, and the estimates for P97 were higher than those for P17, even though, in reality, the model number had nothing to do with the product's quality, novelty, or price.

Anchoring bias has been discussed extensively in psychology (Lieder et al., 2018), medical science (Richie and Josephson, 2018; Pines and Strong, 2019), financial studies (Jetter and Walker, 2017; Shin and Park, 2018), marketing (Esch et al., 2009), organizational studies (Thorsteinson et al., 2008), project management (Lorko et al., 2019), tourism management (Wattanacharoensil and Laornual, 2019), social science (Meub and Proeger, 2015), decision support systems (George et al., 2000), and other fields. For more information about various applications of anchoring bias, see Furnham and Boo (2011).

2.2. *Main causes of anchoring bias and debiasing strategies*

The anchoring and adjustment process is the earliest anchoring bias mechanism (Tversky and Kahneman, 1974), which is known as the "standard paradigm" of anchoring bias in the literature (see e.g. in Section 2.1). In this mechanism, DMs adjust their estimation from the initial value toward the range of plausible values. This adjustment stops in the upper or lower bounds of the plausible values range, which means the adjustment is insufficient (Strack and Mussweiler, 1997). However, this adjustment process is not required for all types of anchors, and researchers have found that it is required when the anchor is self-generated (Furnham and Boo, 2011).

Researchers have mentioned other mechanisms for this bias in recent years, known as "selective accessibility" or "confirmatory search" (Chapman and Johnson, 1994; Strack and Mussweiler, 1997; Mussweiler and Strack, 1999) and "attitude change" (Wegener et al., 2001), and argue that these mechanisms are the best explanations for the anchoring and adjustment process in cases where there is an externally provided anchor.

In the selective accessibility mechanism, DMs test the hypothesis that the externally provided anchor is the correct answer for the decision-making problem. In this way, they look for information that is similar and consistent with the anchor to confirm this hypothesis and ignore information that leads them to reject the hypothesis. The value of the final estimation is affected by this accessibility of information. For example, Mussweiler and Strack (1999) asked subjects in an experiment "whether the average price for a new car is higher or lower than 40,000 Deutschmarks" as a high anchor question. The results showed that the participants named expensive car brands, like BMW, more quickly than the less expensive ones, like Volkswagen Golf, because the expensive brands are more consistent than the cheaper ones with the anchor being provided, which acts as the hypothesis for the DM. However, when subjects in a separate experiment were asked "whether the average price for a new car is higher or lower than 20,000 Deutschmarks," as a low anchor question, the participants named the cheaper brands before they mentioned the more expensive ones.

In the attitude change mechanism, like the elaboration likelihood model (Petty and Cacioppo, 1986), factors like the credibility of the source or the mood of a message recipient act as a persuasion driver and can take on different roles, which can ultimately affect their attitudes in rational or irrational, thoughtful or non-thoughtful ways. In this mechanism, anchors play two different roles as a cue (hint) to provide plausible answer value or by indirectly influencing the DMs toward the anchor and bias the decision-making so that anchor-consistent information is activated. The former role is "low-elaboration" or "non-thoughtful," while the latter role is "high elaborative" or "thoughtful" anchoring.

Researchers have shown that anchoring bias is a robust and pervasive cognitive bias (Furnham and Boo, 2011), and they have tried to provide ways to mitigate the effect of the bias on the final answer of DMs. Table 1 describes these debiasing solutions mentioned in behavioral research. Note that there is an inconsistency between studies examining the effectiveness of these solutions in regard to mitigating the impact of anchoring bias. The solutions included in this table have been proven in experimental studies with their own conditions, and there is no guarantee they can be applied effectively to other problems and situations.

As we can see from Table 1, most debiasing strategies focus on the characteristics of the DM (e.g., expertise, cognitive ability), while some (e.g., "consider-the-opposite" strategy) are related to the decision-making procedure (method). While the former category has been studied in behavioral psychology, in the field of MADM, in this study, we are more interested to see how we could use debiasing strategies as a tool (e.g., "consider-the-opposite" strategy) in devising a method that is less prone to anchoring bias. Based on the literature listed above, this debiasing solution is appropriate for externally provided anchors, and as we know, in the methods considered in this study (SMART and Swing), external anchors (low anchor 10 for SMART and high anchor 100 for Swing) are provided to the subjects involved.

2.3. Anchoring bias in MADM

Despite the vast body of literature on anchoring bias in various research areas, only a few studies have thus far looked at anchoring bias within the context of MADM weighting methods, some of which are briefly discussed below. Rezaei (2021) examined anchoring bias in SMART and Swing methods, compared the estimates provided by the subjects to the actual results of the normative decision-making problem, and showed that potential anchoring bias exists in each of the two methods. Rezaei argued that both methods produce greater (than actual) weights for the less important attributes and lower (than actual) weights for the more important attributes. Generalization of the findings of his study should be made carefully due to two important facts: (i) in his study,

Table 1
Debiasing solutions for anchoring bias

Solution	Description	References
Expertise	The decision-makers (DMs) with a high level of expertise and knowledge about the decision-making problem are less affected by the anchoring bias	Downen et al. (2019), Kaustia et al. (2008), Smith and Windschitl (2015), Welsh et al. (2014), and Wilson et al. (1996)
Incentives	Incentives reduce the impact of anchoring bias because of the accuracy of motivation and increase the adjustment	Epley and Gilovich (2005, 2006), Meub and Proeger (2016), Welsh et al. (2014), and Wright and Anderson (1989)
Personality	Some personality dimensions are more susceptible to anchoring bias than others	Caputo (2014), Eroglu and Croxton (2010), and McElroy and Dowd (2007)
Cognitive ability	Higher cognitive abilities lead to a reduction in the effects of anchoring bias	Bergman et al. (2010) and Meub and Proeger (2016)
Mood	DMs who are sad are more prone to anchoring bias	Bodenhausen et al. (2000), Englich and Soder (2009), and Estrada et al. (1997)
Time pressure	Time pressure increases the likelihood that the DMs fail to make adequate adjustments and increases the impact of anchoring bias	Yik et al. (2019)
Training	Training and being provided with information about anchoring bias and debiasing solutions can reduce the effect of anchoring bias	Adame (2016), Lee et al. (2016), and Meub and Proeger (2016)
Consider-the-opposite	Consider-the-opposite strategy mitigates the effect of anchoring bias (especially for external anchors) because of the multiple and inconsistent anchors being provided	Lord et al. (1984) and Mussweiler et al. (2000)
Group decision-making	Group decision-making leads to a reduction in anchoring bias because of the different anchors considered by different DMs involved	de Wilde et al. (2018), Meub and Proeger (2018), and Sniezek (1992)

graphical representation information is used to check the anchoring bias in subjects' estimation. Existing literature shows that different graphical representation information could lead to different levels of precision in estimation, something that has not been considered in that study and has been extensively studied in the literature (Korhonen and Wallenius, 2008; Gettinger et al., 2013; Liu et al., 2014; Miettinen, 2014; Wachowicz et al., 2019). Regarding the impact of graphical representation information on anchoring bias in decision-making, we refer to Cho et al. (2017). (ii) While in his study, a particular problem has been used for the experimental analysis such that we are able to find the actual/true values as a benchmark to measure the degree of anchoring bias, in almost all real-world decision-making problems, the actual/true weights are unknown (Weber and Borcherding, 1993). In fact, "there is no golden standard for weighting, that is, no measure of a "true" weight is available" (van Til et al., 2014). Lahtinen et al. (2020) looked at how we could reduce anchoring bias in the *even swaps* process. They developed four debiasing methods, including "introducing a virtual reference alternative in the decision problem," "introducing an auxiliary measuring stick attribute," "rotating the reference point," and "restarting the decision process at an intermediate step with a reduced set of alternatives." They described that rotating the reference point is similar to using multiple anchor points when estimating results and argued that these methods could be utilized in weight elicitation using the Swing and trade-off methods to reduce cognitive biases. Montibeller and von Winterfeldt (2018) examined individual and group biases in value and uncertainty judgments in a comprehensive literature review and argued that anchoring bias is a relevant individual cognitive bias in the elicitation of value or utility functions task of MADM and that debiasing strategies try to avoid anchors, providing multiple and counter-anchors and using different experts who use different anchors. Montibeller (2018) looked at behavioral challenges in policy analysis with conflicting objectives by describing his experience in various decision-making projects, for instance, the World Health Organization. The author argued that anchoring bias exists in the modeling values task and proposed debiasing strategies using counterfactuals and multiple experts, among other things. Montibeller and von Winterfeldt (2015b) reviewed biases and debiasing in MADM and proposed some debiasing strategies, including avoiding anchors, providing multiple and counter-anchors, and using different experts with different anchors to mitigate anchoring bias. Buchanan and Corner (1997) examined the performance of the MADM solution methods from a behavioral perspective, describing an experiment involving the effects of the anchoring and adjustment bias in two different interactive solution methods, the "free search interactive" and "Zionts and Wallenius" methods, and found that both the final result and the intermediate iteration solution are likely to be influenced by the starting solution and the solution provided in the previous iteration. Based on that, they hypothesized that the more structured interactive solution method would enhance the effect of anchoring and adjustment bias, and the production scheduling decision problem adapted from Wallenius (1975) was used to examine that hypothesis. Their measure of anchoring relies on a Euclidian distance measure that measures how far participants have moved from their starting solution, determined by rank order weights by the SMART method. The results suggest that subjects are anchored by the starting point in the Zionts and Wallenius method but that the effect of anchoring is not significant in the free search method. Korhonen and Wallenius (1997) reviewed behavioral issues in MADM to improve the success of decision tools in practice, arguing that both in single DM and group DM tasks, the DM's most preferred solution may depend on the starting point and/or the path leading to the most preferred solution. The authors mentioned that there is evidence to suggest that the "path" or sequence in which solutions or settlements are presented to DMs may affect the final choice. Based on those results, the authors suggested looking at the problem from different perspectives and using multiple representations and multiple starting points to reduce anchoring bias in MADM methods.

The literature review presented above revealed that, despite the importance of anchoring bias in MADM weighting, there are gaps in this area that this study tries to cover. One of the main gaps is the lack of experimental research into anchoring bias using a real-life MADM weighting problem, which is also indicated by Rezaei (2021). Another gap involves assessing the effectiveness of existing debiasing tools as argued by Montibeller (2018) and Montibeller and von Winterfeldt (2015a, 2015b, 2018). Finally, different MADM methods could result in different weights for the same problem (Doyle et al., 1997; Bottomley and Doyle, 2001; van Til et al., 2014), and based on Corner and Buchanan (1997), Fox and Clemen (2005); Marttunen et al. (2018), some weighting methods may be more prone to cognitive biases than others, which means it would be interesting to see whether the different methods lead to systematic differences in the estimated preferences. As such, in this research, we examine some weight elicitation methods that use opposite procedures in obtaining a starting point from a DM in a decision-making problem to examine the occurrence of anchoring bias and provide some suggestions to mitigate this bias.

3. MADM methods

There are several methods for determining attribute weights in MADM literature (for more information, see Weber and Borcherding, 1993; Triantaphyllou, 2000; Pöyhönen and Hämäläinen, 2001; Riabacke et al., 2012; Asgharizadeh et al., 2019). In this study, to examine anchoring bias, we looked at three methods (SMART and Swing are tested for anchoring bias, while BWM is used as a possible debiasing method). Below, the three methods are briefly described, after which the reasons for using them are explained in detail. For all methods, a DM evaluates a set of attributes $\{c_1, c_2, \ldots, c_n\}$.

SMART (Edwards, 1977): In this method, the DM starts by ranking the attributes involved in the order of importance, after which the least important attribute is assigned a value of 10 by the DM, and the DM assigns greater values to the other attributes in order of their relative importance. Finally, the weights are calculated by normalizing the values (Equation 1). For a decision-making problem with n attributes ($j = 1, 2, \ldots, n$), s_j is the value that DM assigns to attribute j, and w_j is the importance weight of attribute j.

$$w_j = \frac{s_j}{\sum_{j=1}^{n} s_j}, \ \forall j. \tag{1}$$

Swing (von Winterfeldt and Edwards, 1986): A DM starts from a hypothetical worst alternative scenario, in which all attributes are set to their worst possible levels. Next, the DM is asked to identify which attribute they would prefer most to change from its worst performance level to its best, and the attribute in question is then assigned a value of 100 by the DM, who then repeats this process and assigns values less than or equal to 100 until the worst attribute is assigned a value. The final weight of the attribute j is elicited by normalizing the values using Equation (1).

BWM (Rezaei, 2015, 2016): In this method, the DM first determines the best and worst attributes and, using a scale from 1 to 9 (where 1 represents an equal preference between the attributes and 9 an extreme preference between them), makes a comparison between best attribute B over all the other attributes (a_{Bj}). This will result in vector best-to-others $A_B = (a_{B1}, a_{B2}, \ldots, a_{Bn})$. After that, using the same scale, the DM makes a comparison between all the other attributes and the worst attribute W, (a_{jW}), which will result in vector others-to-worst $A_W = (a_{1W}, a_{2W}, \ldots, a_{nW})^T$. The optimal weight of the attributes is calculated by solving different optimization models. In this study, we use the linear model, which is presented as follows (Rezaei, 2016).

$$\begin{aligned} \min \quad & \xi^L \\ \text{s.t.} \quad & \left|w_B - a_{Bj} w_j\right| \le \xi^L, \ \forall j \\ & \left|w_j - a_{jw} w_W\right| \le \xi^L, \ \forall j. \\ & \sum_j w_j = 1 \\ & w_j \ge 0, \quad \forall j \end{aligned} \tag{2}$$

We chose SMART and Swing to examine anchoring bias because they use opposite procedures in evaluating attributes (SMART starts with a low anchor, and Swing starts with a high anchor). The BWM uses two evaluation vectors, one of which is a comparison between the best attribute (high anchor) and other attributes, and the other one is a comparison between the other attributes and the worst attribute (low anchor). Having both anchors in a single optimization model makes it an excellent candidate for debiasing anchoring bias. We decided to use the linear BWM (Rezaei, 2016) because it generates a unique solution (compared to non-linear BWM (Rezaei, 2015) that may result in multiple optimal solutions) to ensure that the comparison is fair, in light of the fact that the other two methods (SMART and Swing) provide unique solutions.

4. Hypotheses and design of the experiment

4.1. Hypotheses

As stated in Sections 1 and 2, anchoring bias occurs when a DM estimates a numerical value based on the first piece of information (anchor) to which they are exposed, which provides inaccurate estimation values. Several researchers have argued that different MADM weighting methods may lead to different weights for the same attributes (Weber and Borcherding, 1993; Doyle et al., 1997; Bottomley and Doyle, 2001; Pöyhönen and Hämäläinen, 2001; Abel et al., 2020) and that different response scales of the MADM weighting methods lead to different weights for the attributes (Pöyhönen and Hämäläinen, 2001; Pöyhönen et al., 2001). Based on their experimental study on MADM weighting methods, Pöyhönen et al. (2001) concluded that the subjects used a limited set of scores from SMART and Swing methods scale in their evaluations. Their results show that in the case of SMART, only 4% of subjects used scores higher than 100, while in the case of the Swing method, only 2% used scores below 10. In addition, only 18% and 7% of the subjects used scores from all of the ranges available for the SMART and Swing, respectively. According to the authors, as a result, these methods produce different weights. Pöyhönen and Hämäläinen (2001) made similar observations, showing that, for example, for a problem with three attributes, a subject used scores of 100, 90, and 70 with Swing and assigned scores of 40, 20, and 10 with SMART for the same attributes. It would appear that in SMART, where the scoring starts with 10, subjects are more likely to assign the next higher scores closer to 10, while in the case of Swing, where scoring starts with 100, subjects are more likely to assign values closer to 100 for the less important attributes. The same attributes can be assigned different scores, depending on the method being used.

Next, we discuss an example to see how the two methods could lead to different weights. It is important to mention that this example is similar to most of the cases seen in earlier studies (as well as in the present study).

Suppose we ask a subject to use SMART and Swing for a given set of attributes $\{A, B, C, D\}$. The subject, using SMART, identifies the least important attribute as A, followed by B, C, and finally D, as the most important one, and then assigns 10 to A, 20 to B, 35 to C, and 70 to D. The same subject, using Swing and assigning the same order of importance to the attributes, assigns 100 to D, 80 to C, 70 to B and 50 to A. Below, we calculate the weights of the attributes based on the two methods.

The sum of scores for SMART being 135, and for Swing 300, using Equation (1), we can find the weights as follows:

$$\text{SMART}: \; w_A = \frac{10}{135} = 0.074; \; w_B = \frac{20}{135} = 0.148; \; w_C = \frac{35}{135} = 0.260; \; w_D = \frac{70}{135} = 0.518;$$

$$\text{Swing}: \; w_A = \frac{50}{300} = 0.167; \; w_B = \frac{70}{300} = 0.233; \; w_C = \frac{80}{300} = 0.267; \; w_D = \frac{100}{300} = 0.333.$$

As we can see, assigning scores close to the initial point (10 for SMART and 100 for Swing) could lead to SMART assigning lower weight than Swing to the less important attributes A ($0.074 < 0.167$), B ($0.148 < 0.233$), and C ($0.260 < 0.267$), while for the most important attribute D, with SMART, we get a greater weight than with Swing ($0.518 > 0.333$). Rezaei (2021) investigated anchoring bias in the SMART and Swing methods and found similar results. He found that for the smallest alternative (comparable to the least important attribute in our study), SMART produces a lower value than Swing (mean difference: 0.0222), while for the largest alternative (comparable to the most important attribute in our study), the opposite happens, that is, SMART produces a higher value than Swing (mean difference: 0.0239).

In a formal way, in the following proposition, we show how such behavior in the scoring phase of SMART and Swing could affect the normalized scores (weights).

Proposition 1. In an ascendingly ordered set of attributes J, there exists an attribute p such that for $j \leq p$, the normalized scores (weights) of Swing are greater than the normalized scores of SMART, and for $j > p$, the normalized scores of Swing are smaller than the normalized scores of SMART.

Proof. *Suppose that the true score of a DM for attribute j is s_j. In SMART, the stated scores are expected to be lower than their corresponding true values. That is, the stated score is $k_j s_j$, with a multiplier $0 < k_j \leq 1$, and as we move from the least important attribute to the most important one, the multiplier becomes larger, that is, $k_{j+1} \geq k_j$. In the case of Swing, the scores are biased in the opposite direction. That is, the stated score is $l_j s_j$, with a multiplier $l_j \geq 1$,, and as we move from the most important attribute to the least important attribute, the multiplier becomes larger, that is, $l_{j+1} \leq l_j$.*

Following the true, SMART, and Swing scores, the normalized scores (weights) can be found respectively using $\frac{s_j}{\sum_{j=1}^n s_j}$, $\frac{k_j s_j}{\sum_{j=1}^n k_j s_j}$, and $\frac{l_j s_j}{\sum_{j=1}^n l_j s_j}$. It is clear that $\sum_{j=1}^n l_j s_j \geq \sum_{j=1}^n k_j s_j$ or $\sum_{j=1}^n l_j s_j = \theta \sum_{j=1}^n k_j s_j$. Then, for $\frac{l_j}{k_j} \geq \theta$, we have $\frac{k_j s_j}{\sum_{j=1}^n k_j s_j} \leq \frac{l_j s_j}{\sum_{j=1}^n l_j s_j}$, and for $\frac{l_j}{k_j} < \theta$, we have $\frac{k_j s_j}{\sum_{j=1}^n k_j s_j} > \frac{l_j s_j}{\sum_{j=1}^n l_j s_j}$. As k_j belongs to an ascending set and l_j belongs to a descending set, $\frac{l_j}{k_j}$ is decreasing over the ascendingly ordered set of J, where θ is associated with attribute p in this set.

Thus, we complete the proof of Proposition 1.

Similar to Proposition 1, Rezaei (2021) proves that the biased scores of SMART and Swing lead to weights for the less important attributes being higher than their corresponding true ones, while the weights of the more important attributes are lower than their true corresponding ones. Considering his findings and Proposition 1 together, we could also conclude that the range of the weights (the difference between the largest weight and the smallest weight) found by Swing is smaller than the range of the weights found by SMART, and both ranges are smaller than the range of the weights based on their true (unbiased) weights.

Because of these arguments, we want to test the following hypothesis:

H1. SMART weighting, compared to Swing weighting, leads to a smaller weight for the less important attributes and a greater weight for the more important attributes.

Such behavior might not necessarily be seen for all individuals, and our interest is to test this for a sample of subjects.

Rezaei (2021) described the effect of anchoring bias resulting from the response scale of the SMART and Swing weighting methods on the normalized attribute weights and argued and proved that both SMART and Swing methods yield greater weights (compared to their actual values) for the less important attributes and lower weights (compared to their actual values) for the more important attributes. Although he has used an estimation problem for his experimental analyses, the propositions formulated in his paper are not limited to estimation problems. That is, *if* we accept that the starting point of SMART and Swing leads to anchoring bias in scoring the attributes, *then* we can mathematically show that the weights produced by both methods lead to higher weights for the less important attributes and lower weights for the more important attributes. The premise of this argument (having anchoring bias in the scoring of SMART and Swing) has been shown valid in many studies discussed above, so we cannot refute the conclusion, as it has been mathematically proven previously (see Rezaei, 2021). Researchers have also stated that one of the main solutions for reducing anchoring bias in MADM is using a "consider-the-opposite strategy" or providing multiple and counter-anchors (Korhonen and Wallenius, 1997; Montibeller and von Winterfeldt, 2015a, 2015b, 2018; Rezaei, 2021), in other words, considering alternative and contradictive approaches to the problem that are inconsistent with the initial perspective. We think that the BWM, a novel weighting method (Rezaei, 2015, 2016), incorporates this strategy into its procedure. It uses two evaluation vectors, "best-to-others" and "others-to-worst." A DM compares the most important attribute to all the other attributes using the first vector and then compares all the other attributes to the least important attribute using the second vector. The two-vector procedure inherent in this method could act as a "consider-the-opposite strategy" and might cancel out the anchoring bias found in other methods that use one anchor, such as SMART (low anchor) and Swing (high anchor). Recently, Rezaei (2022) showed how the two-vector mechanism of the BWM actually cancels out the effect of anchoring bias that occurs in every single vector. We would argue that because of the anchoring debiasing strategy inherent in the BWM, the final weights of attributes obtained by the BWM should be less prone to anchoring bias, which implies that the BWM, compared to SMART and Swing, should be able to assign lower weights to the less important attributes and higher weights to the more important attributes. That is why we want to test the following hypothesis:

H2. Compared to SMART and Swing, the BWM assigns lower weights to the less important attributes and higher weights to the more important attributes.

Here, we would also like to note that such behavior might not be necessarily seen for all individuals, and our interest is to test this for a sample of subjects.

Table 2
Attributes, sub-attributes, and the decision matrix of the research problem

Attribute	Sub-attribute	Sub-attribute description	Alternatives			
			BRT	*Bus*	*Taxi*	*Metro*
Cost (C1)	Travel cost (C1-1)	Total payment for travel from origin to final destination (toman)	1000	1500	3500	1000
Time (C2)	Travel time (C2-1)	The total time elapsed from the time the vehicle began to move until it reached its destination (minutes)	50	68	60	45
	Waiting time (C2-2)	The total waiting time of the person at the station before the arrival and movement of the vehicle (minutes)	5	5	10	5
	Reliability and punctuality of vehicles mode runs come on schedule to the destination (C2-3)	Non-time deviation of reaching the destination according to the pre-determined or expected plan for that vehicle	High	Medium	low	High
Environment friendly (C3)	Pollution (C3-1)	The amount of air pollution emitted by the vehicle	Low	Medium	High	Very low
Comfort (C4)	The passenger density in the vehicle (C4-1)	Population and congestion within the vehicle	Very high	High	Very low	Very high
	Ease of accessibility to vehicle stop station (C4-2)	The ease and short distance to achieve the desired means of transportation	Low	Very high	Very high	Very low
	Air condition and other equipment in the vehicles (C4-3)	Existence, use, and effectiveness of heating and cooling facilities in the vehicle	High	Medium	High	High

4.2. Design of the experiment

To test the hypotheses described in Section 4.1, we need a MADM problem to obtain data from subjects. We selected *"weighting the attributes and sub-attributes of the evaluation and selection of intra-city public transportation mode in Tehran"* as our test problem. The scenario we presented to each of the subjects was the following.

A respondent has four transport modes (*Bus Rapid Transit (BRT), Bus, Taxi, and Metro*) to move from a fixed point of origin to a fixed destination in the city (a map with all details is provided to the subjects). The four modes are characterized by different attributes as reported in Table 2 (the

Table 3
Subjects' characteristics ($n = 146$)

Characteristics	Levels	Number (percent)
Education level	Master's student	20 (13.7%)
	Master's	65 (44.5%)
	Ph.D. student	61 (41.8%)
Major	Management and industrial engineering	141 (96.6%)
	Miscellaneous (e.g., computer engineering, accounting)	5 (3.4%)
Age	[23, 27)	21 (14.4%)
	[27, 31)	61 (41.8%)
	≥ 31	64 (43.8%)
Gender	Male	81 (55.5%)
	Female	65 (44.5%)

subjects were also given these details). Next, we ask the subjects to evaluate the attributes for their transport mode choice decision-making problem using the three methods (SMART, Swing, and BWM) (see the Appendix for some more details).

The subjects for our experiment were university students in the city of Tehran in Iran who were familiar with MADM methods, a type of subject that is common in this research area (Buchanan and Corner, 1997; Hämäläinen and Alaja, 2008; Rezaei, 2021). In all, 146 subjects took part in our experiment (the characteristics of the subjects are shown in Table 3). The minimum acceptable sample size was checked with GPOWER 3.1[1] (2020) software, and in all cases, our sample size was larger than the size the software provided. We must mention that to enhance the reliability of the results, the subjects' data (weights) with the similar most and least important attribute/sub-attribute in the SMART and Swing methods used to test H1 and the subjects' data with the similar most and least important attribute/sub-attribute in the SMART, Swing, and BWM used to test H2.

The Gorilla platform (https://gorilla.sc) was used for data collection as a novel, powerful, flexible, and user-friendly virtual platform for experimental research (Anwyl-Irvine et al., 2020).

In this study, we considered the weighting methods (SMART, Swing, and BWM) as the experiment factor. The MADM weighting methods used in this research are expert-oriented, which meant that the design of the experiment should cope properly with parameters affecting the subject's preferences, which varied from one subject to another, such as knowledge, personality, thinking style, and cognitive ability, which are mentioned as debiasing solutions in Table 1, and controlled them to examine anchoring bias and the effectiveness of the "consider-the-opposite strategy" as a debiasing solution. Hence, the within-subject design is suitable for the aim of this experiment (Vegas et al., 2015). In this design, the subjects took the experiment's tasks in a randomized order to minimize the carry-over effect. That is, each subject answered the three methods where the order of methods was randomized across the subjects. We used a counter-balancing method for randomization ensuring an almost equal number of possible order combinations.

[1]The most usual, easy to use, effective and efficient software for estimating the required sample size based on each statistical tests (Faul et al., 2007, 2009).

5. Data analysis and discussion

To test the hypotheses from Section 4.1, the weights of attributes and sub-attributes of the problem were calculated for each subject in each method and on the procedure of the MADM methods described in Section 3, after which all the weights were analyzed by SPSS version 26.0 to test the hypotheses.

The initial results showed that, in the case of the SMART method, approximately 5% of the subjects assigned scores above 100 (for the evaluation of the attributes and sub-attributes), while in the case of the Swing method, only about 2% assigned scores below 10. In addition, in the case of the SMART method, some 55% of subjects assigned scores below 50, while in the case of the Swing method, only about 17% assigned scores below 50. These results were in line with existing literature (see, for instance, Pöyhönen and Hämäläinen (2001); Pöyhönen et al. (2001)), which shows that in the case of SMART, where scoring begins with 10, subjects are more likely to assign the attributes/sub-attributes scores close to 10, while in the cases of Swing, where scoring begins with 100, the subjects are more likely to assign scores close to 100. The initial results already show the anchoring bias at the sample level. In the following, we have a closer look at the data to test the two hypotheses.

H1. SMART weighting, compared to Swing weighting, leads to smaller weights for the less important attributes and greater weights for the more important attributes.

To enhance the study's reliability, the data (weights) with the similar most and least important attributes/sub-attributes in the SMART and Swing methods were used to test this hypothesis. In this way, 108 (for the main attributes), 80 (for the sub-attributes of "time"), and 104 (for the sub-attributes of "comfort") are derived from 146 sets of observations.

To test H1, the paired samples t-test was used to show the differences in the effect of the methods on the weights of the attributes/sub-attributes. For the main attributes level, the results show significant differences between the weights of all four attributes. Based on the results, the SMART method produces greater weights than Swing for the most important attribute and the second most important attribute. The means of the most important attribute's weight are 0.420 and 0.362 for the SMART and Swing methods, respectively. The means of the second important attribute's weight are 0.306 and 0.283 for the SMART and Swing methods, respectively (see Table 4 and Fig. 1).

For the other two main attributes, the sign of the mean difference is negative, showing that SMART leads to lower weights than those found by Swing. The means of the third important attribute's weight are 0.207 and 0.228 for the SMART and Swing methods, respectively, while the means of the least important attribute's weight are 0.128 and 0.068 for the SMART and Swing methods, respectively (see Table 4 and Fig. 1).

For the sub-attributes of time and comfort, we see a similar pattern that we found for the main attributes. That is, the mean weights of the most important sub-attributes of time and comfort elicited by SMART are greater than those of the Swing method, while for the least important sub-attributes of time and comfort, SMART leads to smaller mean weights than those found by Swing. The means of weights of the most important sub-attribute of time are 0.527 and 0.425, and the sub-attributes of comfort are 0.547 and 0.445 for the SMART and Swing methods, respectively. On the other hand, the means of the weights of the least important sub-attribute of time are 0.129 and 0.248 and for comfort 0.125 and 0.230 for the SMART and Swing methods, respectively. The

Table 4
Effect of the simple multi-attribute rating technique (SMART) and Swing methods on the weights of the attributes/sub-attributes

Level	Pair	Mean difference	Std. deviation	Std. error mean (paired differences)	95% Confidence interval for paired difference (lower bound)	95% Confidence interval for paired difference (upper bound)	t	df	Sig. (two-tailed)
Attributes	SMART R1–Swing R1	0.058	0.066	0.006	0.045	0.070	9.147	107	0.000
	SMART R2–Swing R2	0.022	0.039	0.004	0.015	0.030	5.930	107	0.000
	SMART R3–Swing R3	−0.020	0.047	0.004	−0.029	−0.012	−4.566	107	0.000
	SMART R4–Swing R4	−0.060	0.054	0.005	−0.071	−0.050	−11.539	107	0.000
Sub-attributes of time	SMART R1–Swing R1	0.102	0.104	0.012	0.079	0.125	8.755	79	0.000
	SMART R2–Swing R2	0.017	0.063	0.007	0.003	0.031	2.356	79	0.021
	SMART R3–Swing R3	−0.119	0.083	0.009	−0.137	−0.100	−12.856	79	0.000
Sub-attributes of comfort	SMART R1–Swing R1	0.102	0.097	0.010	0.083	0.120	10.690	103	0.000
	SMART R2–Swing R2	0.001	0.065	0.006	−0.0117	0.0136	0.151	103	0.880
	SMART R3–Swing R3	−0.104	0.085	0.008	−0.121	−0.088	−12.466	103	0.000

Note: R1 indicates the attribute with the biggest weight, R2 the second, R3 the third and R4 the fourth (or the smallest) weight.

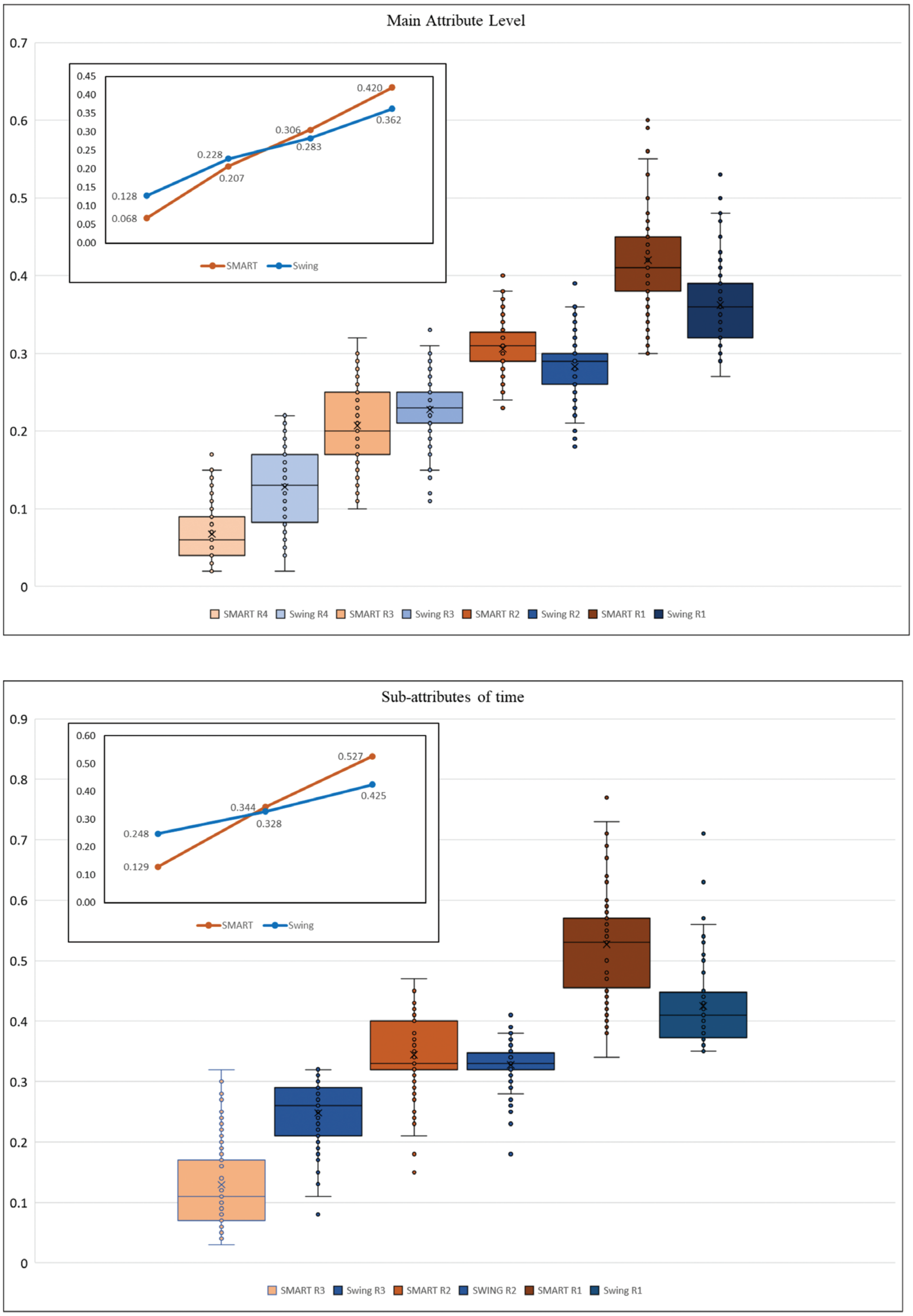

Fig. 1. Weights of the attributes and sub-attributes in simple multi-attribute rating technique (SMART) and Swing (on the top left of each figure, the mean values of the weights and their pattern are summarized).

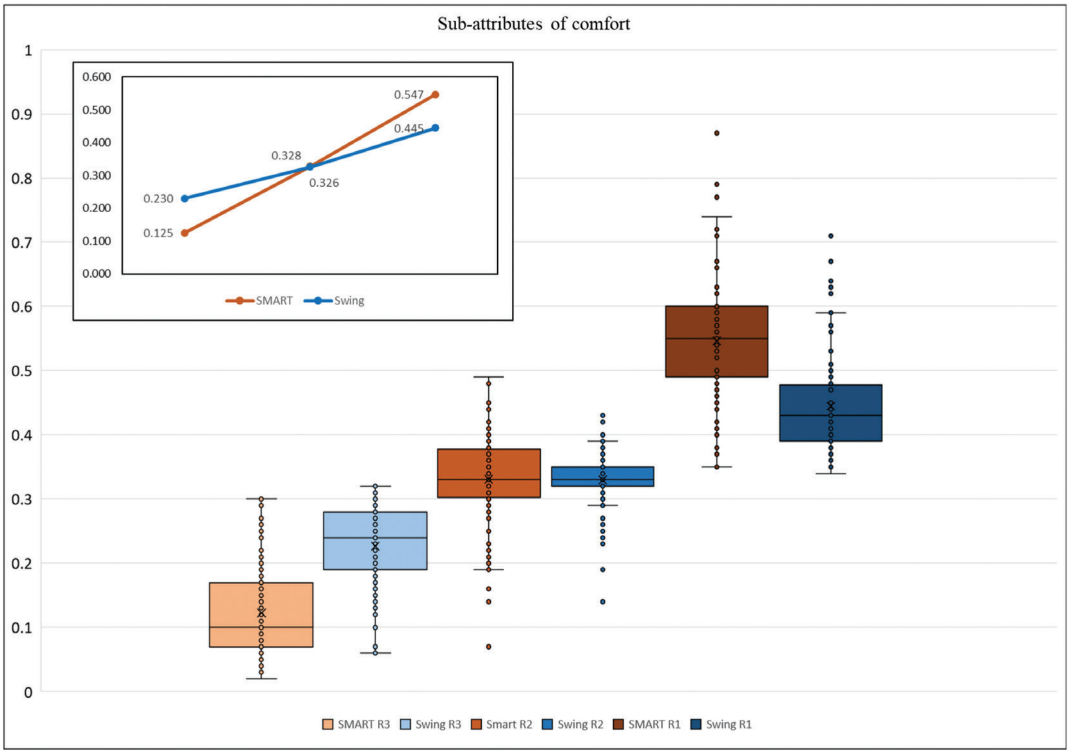

Fig. 1. *Continued.*

second important sub-attribute of time shows a very close mean of weights for both SMART and Swing (0.344 and 0.328, respectively). The same applies to the second important sub-attribute of comfort, where both methods found it very close to each other (for SMART: 0.328, and for Swing: 0.326) and not statistically different (see Table 4 and Fig. 1).

Overall, the results showed that SMART assigns greater weights than Swing to the more important attributes/sub-attributes, while Swing assigns greater weights than SMART to the less important attributes/sub-attributes, which means that H1 is supported. The results are in line with Pöyhönen and Hämäläinen (2001) and Pöyhönen et al. (2001), who argued that different weights for a problem's attributes are the consequence of different response scales of the MADM weighting methods. In the case of SMART, subjects usually assign lower values than in the case of Swing because of the procedure described in Section 3, which, along with the subsequent normalization, the results in greater weights for the more important attributes/sub-attributes and lower weights for the less important attribute/sub-attributes in the case of the SMART method, compared to the Swing method. These results are also consistent with the main findings of Rezaei (2021), which are described in detail in Section 4.1.

H2. Compared to SMART and Swing, the BWM assigns lower weights to the less important attributes and higher weights to the more important attributes.

Similar to H1, the subjects' data (weights), with similar most and least important attributes/sub-attributes in the SMART, Swing, and BWM, were used to test H2 to enhance the study's reliability, using 84 (for the main attributes), 52 (for sub-attributes of "time") and 73 (for sub-attributes of "comfort") from 146 sets of observations, respectively.

To test H2, repeated measures analysis of variance was used to examine the differences between the weights in the three methods. First, for the attributes level, Mauchly's test of sphericity, as an initial testing of assumptions, indicated that the assumption of sphericity had not been met for the effect of the methods on the weights of the main attributes for the most important ($\chi^2 = 7.296$, $p < 0.05$), the second important ($\chi^2 = 35.028$, $p < 0.05$), and the least important attribute ($\chi^2 = 28.568$, $p < 0.05$). The assumption of sphericity is met for the third important attribute ($\chi^2 = 5.694$, $p = 0.058$). Therefore, the Greenhouse–Geisser correction was used to calculate a conservative comparison of the most important, the second important, and the least important main attributes means of the weights. Bonferroni *post hoc* analysis was conducted to determine the methods' weights differentiation at this level. The test of within-subjects effects shows that there was a significant main effect of the methods on the most important (F [1.843, 152.977] $= 261.213$, $p < 0.05$), the second important (F [1.484, 123.178] $=$ 75.985, $p < 0.05$)), the third important (F [2, 166] $= 161.519$, $p < 0.05$)) and the least important (F [1.545, 128.267] $= 84.686$, $p < 0.05$) attributes. As such, to determine the exact differences in the above findings, Bonferroni *post hoc* analyses were conducted. The results show significant differences between the means of the weights for the most important attributes in all three methods. The means of the weights for the most important attribute show that BWM (mean: 0.555) produces greater weights than both SMART (mean: 0.425) and Swing (mean: 0.367). For the least important attributes, we see a significantly different direction. That is, BWM (mean: 0.061) produces smaller weights than SMART (mean: 0.065) and Swing (mean: 0.119) (see Table 5 and Fig. 2).

For the other two middle attributes, we see significant differences between the means of the weights for the second important attribute in all three methods. The means of the weights for the second important attribute show that on average, BWM (mean: 0.240) produces smaller weights than SMART (mean: 0.304) and Swing (mean: 0.292). For the third important attribute, the results show significant differences between the means of the weights in all three methods. The BWM on average (mean: 0.144) produces weights smaller than both SMART (mean: 0.207) and Swing (mean: 0.222) (see Table 5 and Fig. 2).

Similar to the analysis of the attribute level, as far as the sub-attributes of time are concerned, the test of sphericity was met for the most important ($\chi^2 = 2.190$, $p = 0.335$) sub-attribute, and the second important sub-attributes ($\chi^2 = 5.389$, $p = 0.068$), but not for the least important one ($\chi^2 = 6.115$, $p < 0.05$), and the Greenhouse–Geisser correction was used accordingly. The test of within-subjects effects shows that there was a significant main effect of methods on the most important (F [2, 102] $= 109.825$, $p < 0.05$), the second important sub-attribute (F [2, 102] $= 53.782$, $p < 0.05$), and the least important (F[1.794, 91.47] $= 109.546$, $p < 0.05$) sub-attributes. The *post hoc* analysis shows significant differences between all methods means of the weights for the most important sub-attributes. The means of the weights for the most important sub-attribute show that BWM (mean: 0.637) leads to the highest weights, followed by SMART (mean: 0.528) and Swing (mean: 0.421) (see Table 6 and Fig. 2).

The results also show that there are significant differences between the means of the weights for the least important sub-attributes, with the least important sub-attribute showing that BWM

Table 5
Pairwise comparisons of the weights produced by the three methods for the main attributes

Pair	Mean difference	Std. error	Sig.[a]	95% Confidence interval for difference[a] (lower bound)	95% Confidence interval for difference[a] (upper bound)
SMART R1–Swing R1	0.058	0.007	0.000	0.040	0.075
SMART R1–BWM R1	−0.130	0.009	0.000	−0.151	−0.109
Swing R1–BWM R1	−0.187	0.009	0.000	−0.210	−0.164
SMART R2–Swing R2	0.011	0.004	0.006	0.003	0.020
SMART R2–BWM R2	0.064	0.006	0.000	0.048	0.080
Swing R2–BWM R2	0.052	0.006	0.000	0.038	0.067
SMART R3–Swing R3	−0.015	0.005	0.009	−0.026	−0.003
SMART R3–BWM R3	0.063	0.005	0.000	0.051	0.075
Swing R3–BWM R3	0.077	0.004	0.000	0.068	0.087
SMART R4–Swing R4	−0.055	0.006	0.000	−0.068	−0.041
SMART R4–BWM R4	0.004	0.003	0.714	−0.004	0.012
Swing R4–BWM R4	0.059	0.006	0.000	0.045	0.073

Note: R1 indicates the attribute with the biggest weight, R2 the second, R3 the third, and R4 the fourth (or the smallest) weight.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 6
Pairwise comparisons of methods' weights for the sub-attribute of time

Pair	Mean difference	Std. error	Sig.[a]	95% Confidence interval for difference[a] (lower bound)	95% confidence interval for difference[a] (upper bound)
SMART R1–Swing R1	0.108	0.014	0.000	0.073	0.142
SMART R1–BWM R1	−0.109	0.016	0.000	−0.149	−0.069
Swing R1–BWM R1	−0.217	0.014	0.000	−0.250	−0.183
SMART R2–Swing R2	0.014	0.009	0.364	−0.008	0.037
SMART R2–BWM R2	0.097	0.012	0.000	0.068	0.126
Swing R2–BWM R2	0.083	0.009	0.000	0.059	0.106
SMART R3–Swing R3	−0.122	0.011	0.000	−0.150	−0.095
SMART R3–BWM R3	0.012	0.010	0.806	−0.014	0.038
Swing R3–BWM R3	0.134	0.008	0.000	0.114	0.154

Note: R1 indicates the attribute with the biggest weight, R2 the second, and R3 the third (or the smallest) weight.
[a]Adjustment for multiple comparisons: Bonferroni.

(mean: 0.114) leads to the smallest weights, followed by SMART (mean: 0.125) and Swing (mean: 0.248). The difference between SMART and BWM is not statistically significant.

For the middle sub-attribute, the means of weights show that BWM (mean: 0.250) produces smaller weights than Swing (mean: 0.332) and SMART (mean: 0.347). The difference between SMART and Swing is not statistically significant (see Table 6 and Fig. 2).

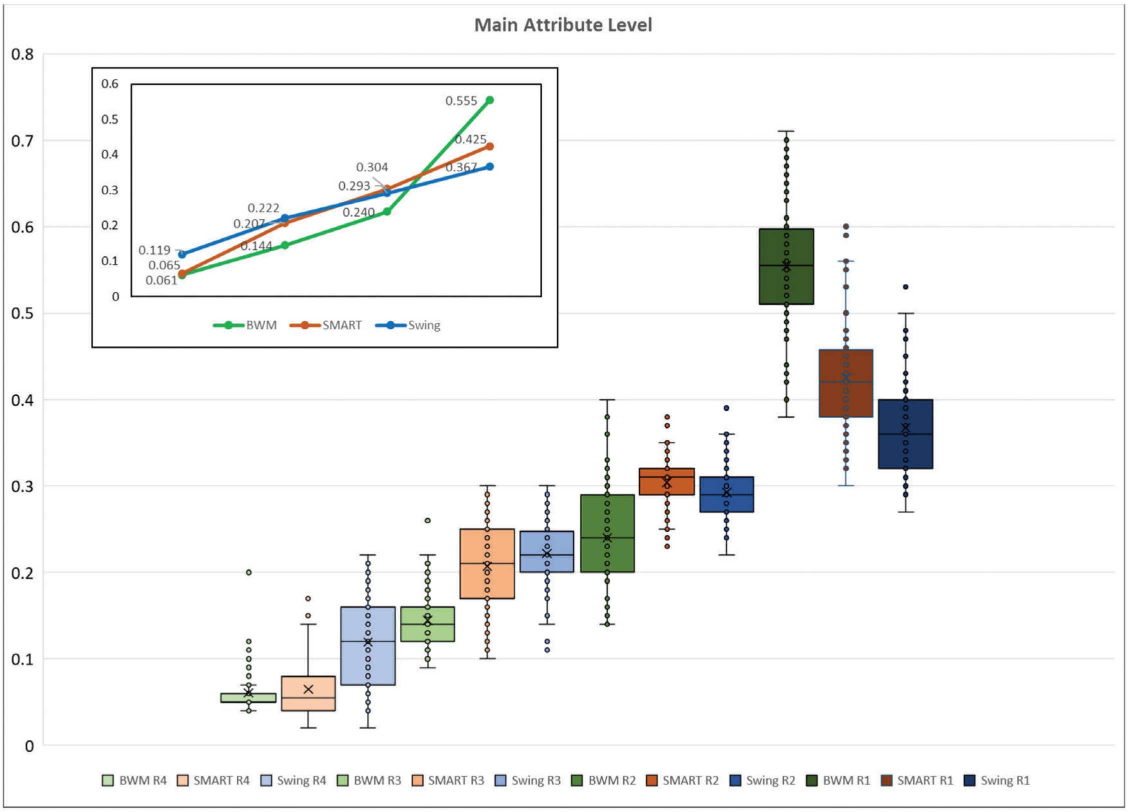

Fig. 2. Weights of the main attributes and sub-attributes in SMART, Swing and best-worst method (BWM) (on the top left of each figure, the mean values of the weights and their pattern are summarized).

These results are similar to those obtained at the main attributes level, in which the BWM assigns greater weights to the more important sub-attributes and lower weights to the less important sub-attributes, compared to SMART and Swing.

Finally, for the sub-attributes of comfort, the sphericity test was met both for the most important ($\chi^2 = 3.263$, $p = 0.196$) and for the least important ($\chi^2 = 5.280$, $p = 0.071$) but not for the second important ($\chi^2 = 17.743$, $p < 0.05$), and the Greenhouse–Geisser correction was used accordingly. The test of within-subjects effects shows that there was a significant main effect of the methods on the most important ($F\ [2, 144] = 148.164$, $p < 0.05$), the second important attribute ($F\ [1.638, 117.924] = 67.679$, $p < 0.05$), and the least important ($F\ [2, 144] = 113.185$, $p < 0.05$) sub-attributes. The results of *post hoc* analysis show that there are significant differences between the means of weights for the most important sub-attributes, with BWM (mean: 0.657), SMART (mean: 0.544), and Swing (mean: 0.443) assigning a higher to smaller weights, respectively (see Table 7 and Fig. 2).

The means of weights for the second important sub-attribute show that BWM (mean: 0.239), Swing (mean: 0.330), and SMART (mean: 0.335) have lower to higher weights, respectively. The difference between SMART and Swing is not statistically significant.

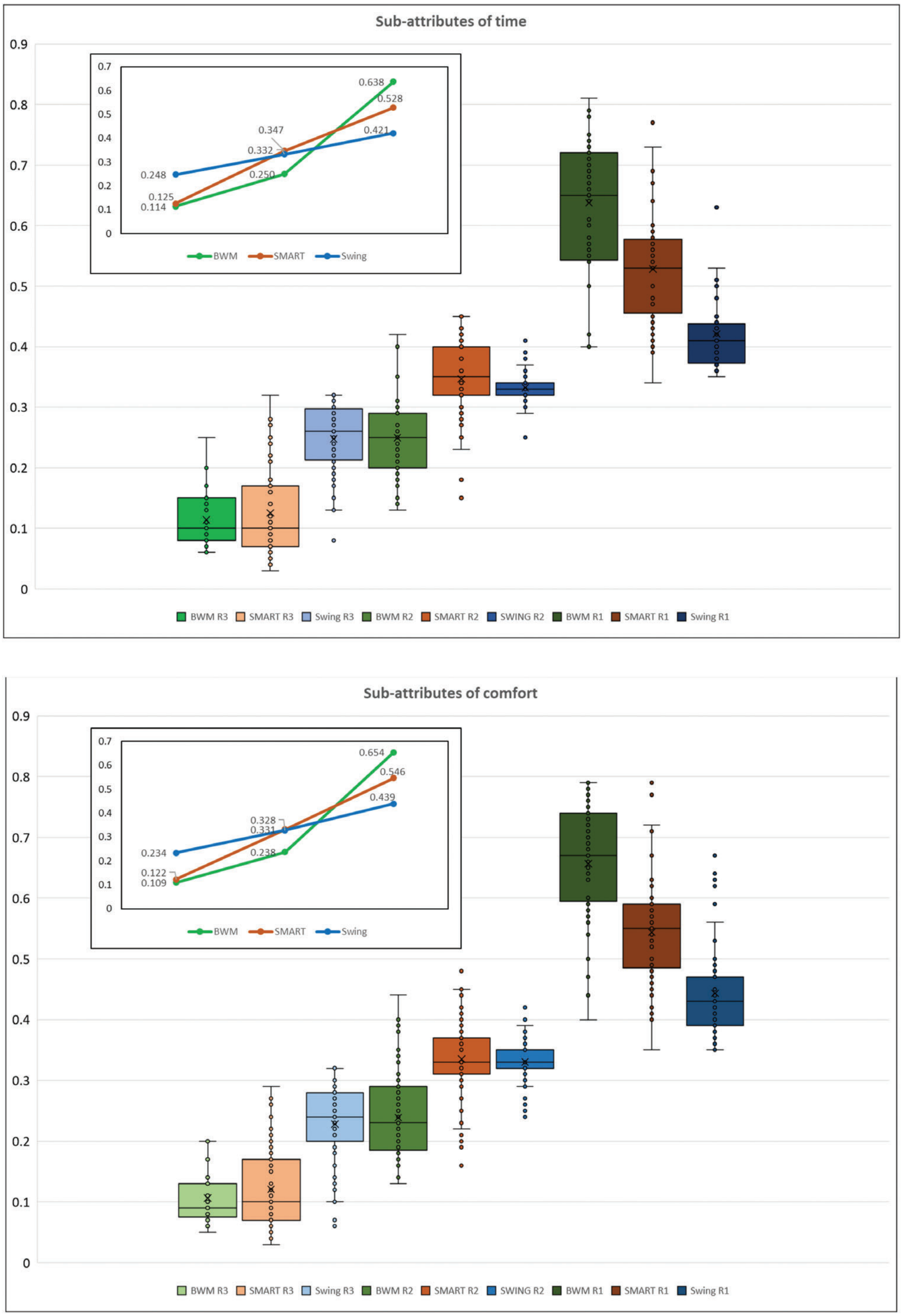

Fig. 2. *Continued.*

Table 7
Pairwise comparisons of methods' weights for the sub-attributes of comfort

Pair	Mean difference	Std. error	Sig.[a]	95% Confidence interval for difference[a] (lower bound)	95% Confidence interval for difference[a] (upper bound)
SMART R1–Swing R1	0.101	0.011	0.000	0.073	0.128
SMART R1–BWM R1	−0.113	0.013	0.000	−0.146	−0.080
Swing R1–BWM R1	−0.214	0.013	0.000	−0.245	−0.183
SMART R2–Swing R2	0.005	0.008	1.000	−0.014	0.024
SMART R2–BWM R2	0.096	0.011	0.000	0.068	0.123
Swing R2–BWM R2	0.091	0.008	0.000	0.070	0.112
SMART R3–Swing R3	−0.107	0.010	0.000	−0.132	−0.083
SMART R3–BWM R3	0.015	0.008	0.233	−0.005	0.034
Swing R3–BWM R3	0.122	0.008	0.000	0.101	0.142

Note: R1 indicates the attribute with the biggest weight, R2 the second, and R3 the third (or the smallest) weight.
[a]Adjustment for multiple comparisons: Bonferroni.

The results also show that there are significant differences between the means of weights for the least important sub-attributes (Table 7), with BWM (mean: 0.106), SMART (mean: 0.121), and Swing (mean: 0.228) assigned lower to higher weights, respectively (the difference between SMART and BWM is not statistically significant). These results are similar to the results regarding the previous attributes and sub-attributes, with BWM assigning lower weight to the less important sub-attributes and higher weights to the more important attributes, compared to the SMART and Swing weighting methods (see Table 7 and Fig. 2).

Overall, the results indicate that the BWM leads to greater weights than SMART and Swing with regard to the more important attributes/sub-attributes and to lower weights for the less important attributes/sub-attributes (Fig. 2), which means that this hypothesis is supported. The results are in line with Montibeller and von Winterfeldt, (2015a, 2015b, 2018), Korhonen and Wallenius (1997), and Rezaei (2021).

5.1. BWM and its mitigation strategy

Several researchers have suggested using multiple anchors for debiasing (Montibeller and von Winterfeldt, 2015b, 2018). Here, we would like to have a closer look at the debiasing mechanism of the BWM. BWM uses two reference points in conducting pairwise comparisons. This is one of the features that makes it different from methods such as SMART and Swing with one reference point. The first reference point (best) could lead to the best and the worst attributes being overweighted, while the other reference point (worst) could lead to the best and the worst attributes being underweighted. For other attributes, their behavior is opposite to each other. That is, the others-to-worst pairwise comparisons, compared to the best-to-others vector, lead to higher weights for the middle attributes. It is important to note that the effect of anchoring bias using these two reference points in the BWM is different than that of SMART and Swing, as in the BWM, we do not assign scores

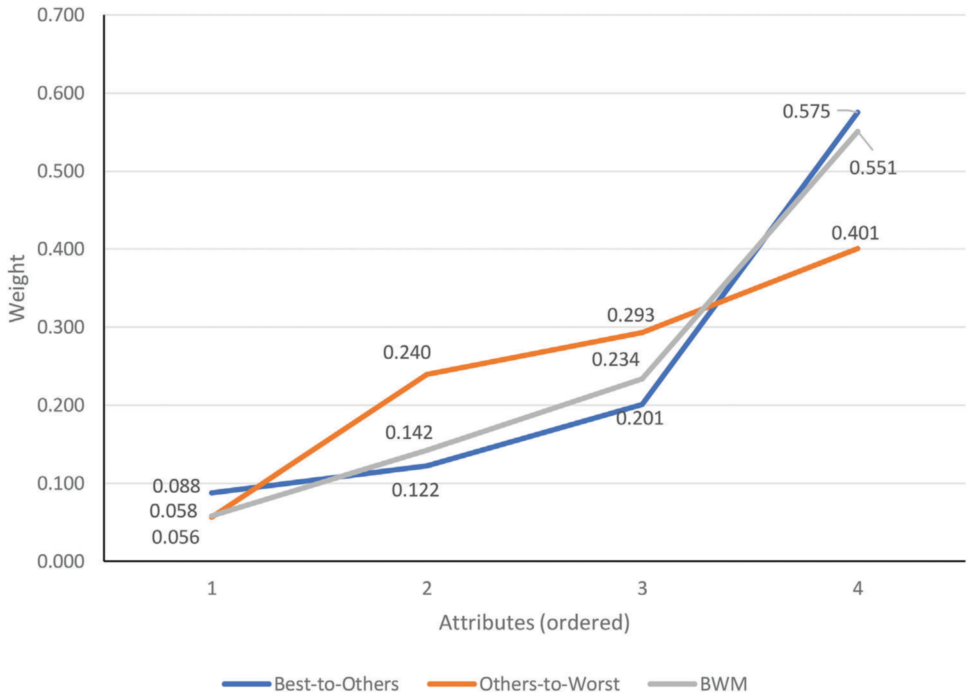

Fig. 3. Weights of the main attributes (ordered) using best-to-others vector, others-to-worst vector, and BWM.

to the attributes but to the pairwise comparisons (for detailed mathematical and numerical support of the anchoring mechanism of these two reference points, see Rezaei, 2022). It is also important to note that producing the two sets of weights based on best-to-others and others-to-worst are only for the purpose of showing their opposite behavior and that the final weights of the BWM are based on this mitigation strategy inherent to the method. Because the two vectors are the input of a single optimization problem in the BWM, the anchoring effects of the two vectors that are in opposite directions are canceled out. In this part, we report the results for the main attributes (the analyses of the other levels are the same).

As seen from Fig. 3, while using a single vector (or a single reference point), the weights of the most important and the least important attributes are either overweighted (using only reference point Best) or underweighted (using only reference point Worst), the results of the BWM show somehow a compromised solution that reflects its anchoring bias mitigation feature.

5.2. Additional support for the hypotheses

As we discussed above, the weights found by SMART and Swing are affected by the biased scores. In both methods, the more important attributes are expected to be underweighted, while the less important attributes are likely to be overweighted. This means that the range of the weights in both methods is expected to be smaller than their true range. Also considering Proposition 1, we can conclude that the range of Swing should be smaller than that of the SMART. We do not have the real weights; nonetheless, we can test the difference between the ranges of Swing, SMART, and BWM, the results of which are as follows. For the main attributes level, the mean range of Swing (0.248, s.d. 0.104) is less than the mean range of SMART (0.360, s.d. 0.079) ($p < 0.05$), and they are smaller than the mean range of the BWM (0.494, s.d. 0.081) ($p < 0.05$). For the sub-attributes of time, the mean range of Swing (0.174, s.d. 0.108) is less than the mean range of SMART (0.403, s.d. 0.156) ($p < 0.05$), and they are smaller than the mean range of the BWM (0.524, s.d. 0.146) ($p < 0.05$). For the sub-attributes of comfort, the mean range of Swing (0.215, s.d. 0.136) is less than the mean range of SMART (0.423, s.d. 0.139) ($p < 0.05$), and they are smaller than the mean range of the BWM (0.551, s.d. 0.128) ($p < 0.05$).

The observation that the mean range of the weights of the attributes for the main attributes and the sub-attributes for the BWM is greater than their associated mean range of the weights of Swing and SMART leads us to conclude that the weights found by the BWM should be less biased than the weights found by SMART and Swing.

6. Conclusion and future research

The weighting of the attributes is one of the most important tasks in MADM. Most MADM weighting methods are based on the judgments of DMs, who rely on a limited number of heuristic principles that reduce the complex tasks involved. Generally, these heuristic principles are beneficial and simplify the decision-making process, but they sometimes lead to errors and deviations from rational decision-making. As a result, subjective judgments designed to assign weights in MADM problems are prone to cognitive biases and risk distorting the optimal results, which means that we need to examine these biases in MADM weighting methods and look for effective debiasing solutions to try and reduce their impact on the final results.

Some studies have investigated cognitive biases in MADM. The aim of our research was to examine anchoring bias, one of the most important biases, by an experimental study using a real-life decision-making problem, for which we selected two well-known MADM methods, SMART and Swing, which have different starting points in their evaluations. The results show that whereas SMART tends to overestimate the weights of the more important (sub)attributes, Swing tends to do the opposite, that is, it underestimates the weight of the less important (sub)attributes. The occurrence of anchoring bias in the scoring phase of SMART and Swing has a natural consequence, meaning that both SMART and Swing underweight the more important attributes, while they overweigh the less important attributes. We then demonstrated that the BWM assigns greater weights than SMART and Swing to the more important (sub)attributes and lower weights to the less important (sub)attributes. One, however, could argue that we do not know if such differences between the weights of BWM and SMART and Swing do not lead to overweighting the more important attributes and underweighting the less important attributes by the BWM. Although we showed the anchoring mitigation strategy inherent in the BWM, further research is needed for stronger support.

In this study, we have not looked at variables like the number and type of attributes, and it is an interesting suggestion for future studies. For instance, in our study, we looked at three and four (sub)attributes for each elicitation task; however, it is interesting to investigate the effect of "number of (sub)-attributes" on anchoring bias. It would also be interesting to research whether and how the type of evaluation (e.g., numerical vs. linguistic) could lead to different degrees of anchoring bias. Based on the anchoring bias mechanism described in Section 2.2, two mechanisms, named "standard paradigm" and "selective accessibility," may be the underlying cause of anchoring bias in the results of MADM weighting methods. In the standard paradigm, subjects adjust their estimation from the initial value (i.e., 10 in SMART or 100 in Swing) toward the range of plausible values. This adjustment stops in the upper or lower bounds of the range of plausible values (i.e., 20, 30, or 40 for SMART and 90, 80, or 70 for Swing), leading to insufficient adjustment and distorting the final weights of the (sub)attributes. Also, in the case of the selective accessibility mechanism, the subjects test the hypothesis that the externally provided anchor (i.e., 10 for SMART or 100 for Swing) is the correct answer for the less or more important attributes/sub-attributes, and because of the executive procedure of the methods that force them to use 10 or 100, they use these values. For the other attributes/sub-attributes, the final value is affected by this accessibility of information inherent in the selective accessibility mechanism and leads to lower values in SMART or higher values in Swing. In that regard, it would be interesting to investigate whether eliminating the scoring limits (i.e., 10 for SMART and 100 for Swing) could mitigate the impact of anchoring bias.

References

Abel, E., Galpin, I., Paton, N.W., Keane, J.A., 2020. Pairwise comparisons or constrained optimization? A usability evaluation of techniques for eliciting decision priorities. *International Transactions in Operational Research 29* 5, 3190–3206.

Adame, B.J., 2016. Training in the mitigation of anchoring bias: a test of the consider-the-opposite strategy. *Learning and Motivation* 53, 36–48.

Anwyl-Irvine, A.L., Massonnié, J., Flitton, A., Kirkham, N., Evershed, J.K., 2020. Gorilla in our midst: an online behavioral experiment builder. *Behavior Research Methods* 52, 1, 388–407.

Asgharizadeh, E., Taghizadeh Yazdi, M., Mohammadi Balani, A., 2019. An output-oriented classification of multiple attribute decision-making techniques based on fuzzy c-means clustering method. *International Transactions in Operational Research* 26, 6, 2476–2493.

Bazerman, M.H., Moore, D.A., 1994. *Judgment in Managerial Decision Making*. Wiley, New York.

Bergman, O., Ellingsen, T., Johannesson, M., Svensson, C., 2010. Anchoring and cognitive ability. *Economics Letters* 107, 1, 66–68.

Bodenhausen, G.V., Gabriel, S., Lineberger, M., 2000. Sadness and susceptibility to judgmental bias: the case of anchoring. *Psychological Science* 11, 4, 320–323.

Borcherding, K., von Winterfeldt, D., 1988. The effect of varying value trees on multiattribute evaluations. *Acta Psychologica* 68, 1-3, 153–170.

Bottomley, P.A., Doyle, J.R., 2001. A comparison of three weight elicitation methods: good, better, and best. *Omega* 29, 6, 553–560.

Buchanan, J.T., Corner, J., 1997. The effects of anchoring in interactive MCDM solution methods. *Computers & Operations Research* 24, 10, 907–918.

Caputo, A., 2014. Relevant information, personality traits and anchoring effect. *International Journal of Management and Decision Making* 13, 1, 62–76.

Chapman, G.B., Johnson, E.J., 1994. The limits of anchoring. *Journal of Behavioral Decision Making* 7, 4, 223–242.

Cho, I., Wesslen, R., Karduni, A., Santhanam, S., Shaikh, S., Dou, W., 2017. The anchoring effect in decision-making with visual analytics. 2017 IEEE Conference on Visual Analytics Science and Technology (VAST), October 3–6, Phoenix, AZ, pp. 116–126.

Corner, J.L., Buchanan, J.T., 1997. Capturing decision maker preference: experimental comparison of decision analysis and MCDM techniques. *European Journal of Operational Research* 98, 1, 85–97.

Critcher, C.R., Gilovich, T., 2008. Incidental environmental anchors. *Journal of Behavioral Decision Making* 21, 3, 241–251.

de Wilde, T.R., Ten Velden, F.S., de Dreu, C.K., 2018. The anchoring bias in groups. *Journal of Experimental Social Psychology* 76, 116–126.

Deniz, N., 2020. Cognitive biases in MCDM methods: an embedded filter proposal through sustainable supplier selection problem. *Journal of Enterprise Information Management* 33, 5, 947–963.

Downen, T., Furner, Z., Cataldi, B., 2019. The effects on anchoring of increasing quantities of disconfirming evidence. *International Journal of Management and Decision Making* 18, 3, 309–331.

Doyle, J.R., Green, R.H., Bottomley, P.A., 1997. Judging relative importance: direct rating and point allocation are not equivalent. *Organizational Behavior and Human Decision Processes* 70, 1, 65–72.

Edwards, W., 1977. How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics* 7, 5, 326–340.

Englich, B., Soder, K., 2009. Moody experts—How mood and expertise influence judgmental anchoring. *Judgment and Decision Making* 4, 1, 41.

Epley, N., Gilovich, T., 2005. When effortful thinking influences judgmental anchoring: differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making* 18, 3, 199–212.

Epley, N., Gilovich, T., 2006. The anchoring-and-adjustment heuristic: why the adjustments are insufficient. *Psychological Science* 17, 4, 311–318.

Eroglu, C., Croxton, K.L., 2010. Biases in judgmental adjustments of statistical forecasts: the role of individual differences. *International Journal of Forecasting* 26, 1, 116–133.

Esch, F.R., Schmitt, B.H., Redler, J., Langner, T., 2009. The brand anchoring effect: a judgment bias resulting from brand awareness and temporary accessibility. *Psychology & Marketing* 26, 4, 383–395.

Estrada, C.A., Isen, A.M., Young, M.J., 1997. Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational Behavior and Human Decision Processes* 72, 1, 117–135.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods* 41, 4, 1149–1160.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2, 175–191.

Fischer, G.W., 1995. Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes* 62(3), 252–266.

Fischer, G. W., Damodaran, N., Laskey, K. B., Lincoln, D., 1987. Preferences for proxy attributes. *Management Science* 33(2), 198–214.

Fox, C.R., Clemen, R.T., 2005. Subjective probability assessment in decision analysis: partition dependence and bias toward the ignorance prior. *Management Science* 51, 9, 1417–1432.

Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics* 40, 1, 35–42.

Gabrielli, W.F. Jr., von Winterfeldt, D., 1978. Are important weights sensitive to the range of alternatives in multiattribute utility measurement. Available at: https://apps.dtic.mil/sti/citations/ADA073366 (accessed 1 March 2020).

George, J.F., Duffy, K., Ahuja, M., 2000. Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems* 29, 2, 195–206.

Gettinger, J., Kiesling, E., Stummer, C., Vetschera, R., 2013. A comparison of representations for discrete multi-criteria decision problems. *Decision Support Systems* 54, 2, 976–985.

Hämäläinen, R.P., Alaja, S., 2008. The threat of weighting biases in environmental decision analysis. *Ecological Economics* 68, 1-2, 556–569.

Jacobi, S.K., Hobbs, B.F., 2007. Quantifying and mitigating the splitting bias and other value tree-induced weighting biases. *Decision Analysis* 4, 4, 194–210.

Jetter, M., Walker, J.K., 2017. Anchoring in financial decision-making: evidence from Jeopardy! *Journal of Economic Behavior & Organization* 141, 164–176.

Kaustia, M., Alho, E., Puttonen, V., 2008. How much does expertise reduce behavioral biases? The case of anchoring effects in stock return estimates. *Financial Management* 37, 3, 391–412.

Korhonen, P., Wallenius, J., 1997. Behavioral issues in MCDM: Neglected research questions. In Climaco, J. (ed.), *Multicriteria Analysis*. Springer, Cham, pp. 412–422.

Korhonen, P., Wallenius, J., 2008. *Visualization in the multiple objective decision-making framework*. In Branke, J., Deb, K., Miettinen, K., Słowiński, R. (eds) *Multiobjective Optimization*. Springer, Cham, pp. 195–212.

Lahtinen, T.J., Hämäläinen, R.P., Jenytin, C., 2020. On preference elicitation processes which mitigate the accumulation of biases in multi-criteria decision analysis. *European Journal of Operational Research* 282, 1, 201–210.

Lee, Y.-H., Dunbar, N.E., Miller, C.H., Lane, B.L., Jensen, M.L., Bessarabova, E., Burgoon, J.K., Adame, B.J., Valacich, J.J., Adame, E.A., Bostwick, E., Piercy, C.W., Elizondo, J., Wilson, S. N., 2016. Training anchoring and representativeness bias mitigation through a digital game. *Simulation & Gaming* 47, 6, 751–779.

Lieder, F., Griffiths, T.L., Huys, Q.J., Goodman, N.D., 2018. The anchoring bias reflects the rational use of cognitive resources. *Psychonomic Bulletin & Review* 25, 1, 322–349.

Lin, S.-W., 2013. An investigation of the range sensitivity of attribute weight in the analytic hierarchy process. *Journal of Modelling in Management* 8, 1, 65–80.

Liu, S., Cui, W., Wu, Y., Liu, M., 2014. A survey on information visualization: recent advances and challenges. *The Visual Computer* 30, 12, 1373–1393.

Lord, C.G., Lepper, M.R., Preston, E., 1984. Considering the opposite: a corrective strategy for social judgment. *Journal of Personality and Social Psychology* 47, 6, 1231–1243.

Lorko, M., Servátka, M., Zhang, L., 2019. Anchoring in project duration estimation. *Journal of Economic Behavior & Organization* 162, 49–65.

Marttunen, M., Belton, V., Lienert, J., 2018. Are objectives hierarchy-related biases observed in practice? A meta-analysis of environmental and energy applications of Multi-Criteria Decision Analysis. *European Journal of Operational Research* 265, 1, 178–194.

McElroy, T., Dowd, K., 2007. Susceptibility to anchoring effects: how openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making* 2, 1, 48.

Meub, L., Proeger, T., 2016. Can anchoring explain biased forecasts? Experimental evidence. *Journal of Behavioral and Experimental Finance* 12, 1–13.

Meub, L., Proeger, T., 2018. Are groups 'less behavioral'? The case of anchoring. *Theory and Decision* 85, 2, 117–150.

Meub, L., Proeger, T.E., 2015. Anchoring in social context. *Journal of Behavioral and Experimental Economics* 55, 29–39.

Miettinen, K., 2014. Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectrum* 36, 1, 3–37.

Montibeller, G., 2018. *Behavioral challenges in policy analysis with conflicting objectives. In Gel, E., Ntaimo, L. (eds)* Recent Advances in Optimization and Modeling of Contemporary Problems. INFORMS, Catonsville, MD, pp. 85–108.

Montibeller, G., von Winterfeldt, D., 2015a. Biases and debiasing in multi-criteria decision analysis. 2015 48th Hawaii International Conference on System Sciences, January 5—8, Kauai, HI, pp. 1218–1226.

Montibeller, G., von Winterfeldt, D., 2015b. Cognitive and motivational biases in decision and risk analysis. *Risk Analysis* 35, 7, 1230–1251.

Montibeller, G., von Winterfeldt, D., 2018. Individual and group biases in value and uncertainty judgments. In Dias, L.C., Morton, A., Quigley, J. (eds), *Elicitation*, Vol. 261. Springer, Cham, pp. 377–392.

Mussweiler, T., Strack, F., 1999. Hypothesis-consistent testing and semantic priming in the anchoring paradigm: a selective accessibility model. *Journal of Experimental Social Psychology* 35, 2, 136–164.

Mussweiler, T., Strack, F., Pfeiffer, T., 2000. Overcoming the inevitable anchoring effect: considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin* 26, 9, 1142–1150.

Petty, R.E., Cacioppo, J.T., 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. Springer-Verlag, New York.

Pines, J.M., Strong, A., 2019. Cognitive biases in emergency physicians: a pilot study. *The Journal of Emergency Medicine* 27, 2, 168–172.

Pöyhönen, M., Hämäläinen, R.P., 1998. Notes on the weighting biases in value trees. *Journal of Behavioral Decision Making* 11, 2, 139–150.

Pöyhönen, M., Hämäläinen, R.P., 2000. There is hope in attribute weighting. *INFOR: Information Systems and Operational Research* 38, 3, 272–282. https://doi.org/10.1080/03155986.2000.11732412

Pöyhönen, M., Hämäläinen, R.P., 2001. On the convergence of multiattribute weighting methods. *European Journal of Operational Research* 129, 3, 569–585.

Pöyhönen, M., Vrolijk, H., Hämäläinen, R.P., 2001. Behavioral and procedural consequences of structural variation in value trees. *European Journal of Operational Research* 134(1), 216–227.

Rezaei, J., 2015. Best-worst multi-criteria decision-making method. *Omega* 53, 49–57.

Rezaei, J., 2016. Best-worst multi-criteria decision-making method: some properties and a linear model. *Omega* 64, 126–130. https://doi.org/10.1016/j.omega.2015.12.001

Rezaei, J., 2021. Anchoring bias in eliciting attribute weights and values in multi-attribute decision making. *Journal of Decision Systems* 30(1), 72–96.

Rezaei, J., 2022. The balancing role of best and worst in best-worst method. In Rezaei, J., Brunelli, M., Mohammadi, M. (eds) *Advances in Best-Worst Method. Lecture Notes in Operations Research*. Springer, Cham, pp. 1–15.

Rezaei, J., Arab, A., Mehregan, M., 2022. Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research. *Journal of Behavioral Decision Making* 35, 2, e2262.

Riabacke, M., Danielson, M., Ekenberg, L., 2012. State-of-the-art prescriptive criteria weight elicitation. *Advances in Decision Sciences* 2012, 1–24.

Richie, M., Josephson, S.A., 2018. Quantifying heuristic bias: anchoring, availability, and representativeness. *Teaching and Learning in Medicine* 30, 1, 67–75.

Shin, H., Park, S., 2018. Do foreign investors mitigate anchoring bias in stock market? Evidence based on post-earnings announcement drift. *Pacific-Basin Finance Journal* 48, 224–240.

Simon, H.A., 1957. *Models of Man; Social and Rational*. Wiley, Oxford, England.

Smith, A.R., Windschitl, P.D., 2015. Resisting anchoring effects: the roles of metric and mapping knowledge. *Memory & Cognition* 43, 7, 1071–1084.

Sniezek, J.A., 1992. Groups under uncertainty: an examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes* 52, 1, 124–155.

Strack, F., Mussweiler, T., 1997. Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. *Journal of Personality and Social Psychology* 73, 3, 437–446.

Thorsteinson, T.J., Breier, J., Atwell, A., Hamilton, C., Privette, M., 2008. Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes* 107, 1, 29–40.

Triantaphyllou, E., 2000. *Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, Cham.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185, 4157, 1124–1131.

Ünveren, B., Baycar, K., 2019. Historical evidence for anchoring bias: the 1875 cadastral survey in Istanbul. *Journal of Economic Psychology* 73, 1–14.

van Til, J., Groothuis-Oudshoorn, C., Lieferink, M., Dolan, J., Goetghebeur, M., 2014. Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework. *Cost Effectiveness and Resource Allocation* 12, 1, 1–11.

Vegas, S., Apa, C., Juristo, N., 2015. Crossover designs in software engineering experiments: benefits and perils. *IEEE Transactions on Software Engineering* 42, 2, 120–135.

von Nitzsch, R., Weber, M., 1993. The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science* 39, 8, 937–943.

von Winterfeldt, D., Edwards, W., 1986. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge.

Wachowicz, T., Kersten, G.E., Roszkowska, E., 2019. How do I tell you what I want? Agent's interpretation of principal's preferences and its impact on understanding the negotiation process and outcomes. *Operational Research* 19, 4, 993–1032.

Wallenius, J., 1975. Comparative evaluation of some interactive approaches to multicriterion optimization. *Management Science* 21, 12, 1387–1396.

Wattanacharoensil, W., La-ornual, D., 2019. A systematic review of cognitive biases in tourist decisions. *Tourism Management* 75, 353–369.

Weber, M., Borcherding, K., 1993. Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research* 67, 1, 1–12.

Weber, M., Eisenführ, F., von Winterfeldt, D., 1988. The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science* 34, 4, 431–445.
Wegener, D.T., Petty, R.E., Detweiler-Bedell, B.T., Jarvis, W.B.G., 2001. Implications of attitude change theories for numerical anchoring: anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology* 37, 1, 62–69.
Welsh, M.B., Delfabbro, P.H., Burns, N.R., Begg, S.H., 2014. Individual differences in anchoring: traits and experience. *Learning and Individual Differences* 29, 131–140.
Wilson, T.D., Houston, C.E., Etling, K.M., Brekke, N., 1996. A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General* 125, 4, 387–402.
Wright, W.F., Anderson, U., 1989. Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes* 44, 1, 68–82.
Yik, M., Wong, K.F.E., Zeng, K.J., 2019. Anchoring and adjustment during affect inferences. *Frontiers in Psychology* 9, 2567.

Appendix

Due to the concurrence of the data collection stage of this study with the prevalence of the coronavirus disease 2019 (COVID-19) outbreak, it was inevitable to change the pre-determined data collection method, which was in the lab of faculty and using MEDIALAB software. For this purpose, after a comprehensive review of virtual platforms and matching their capabilities with research needs, the Gorilla site was finally selected. This platform is one of the new platforms for virtually conducting experimental research and has attracted the attention of many researchers today. Several papers have been published, especially after 2019, using this platform. The main reasons for this choice are a flexible and comprehensive experiment design mechanism, questionnaire design, randomization mechanisms, data completion time storage, comprehensive management of subjects, and an appropriate and easy user interface. Due to the specific application of MADM methods in the experimental research literature, it was impossible to construct particular questionnaires for each method in a pre-prepared manner. Therefore, the HTML programming language was used to overcome this limitation.

After that, the unique experiment link is sent to the email address of the subjects so that each unique link can be used by only one subject with a unique token. First, the problem description is provided to them in detail (see Section 4 of the paper; please note that we also used some tables, examples, and visualizations in the questionnaire). Then, the subjects were randomly assigned (without replacement) in the Gorilla to 3-item questionnaires (three methods: BWM, SMART, and Swing and three levels: attributes, sub-attributes of time, and sub-attributes of comfort), and each subject must do all of these methods and levels. After that, a demography questionnaire was used to collect the characteristics of the subjects. Besides, the possibility of going back to the previous step in experimenting is disabled in all stages. Other settings include no browser restrictions, the geographical location of internet connection and Internet speed, and limiting the subjects' response platform to personal computers and tablets. Also, the descriptions of each stage of the experiment and a numerical example are provided briefly on each page related to that experiment task.

Below, we provide some examples from the questionnaire. Please note that the platform we used has user-friendly options that we incorporated in structuring the questions, which are not shown here.

An example of the SMART (sub-attributes of time)

Considering the goal of the problem (selection of intra-city public transportation mode in the described case), based on your personal preferences, start by ranking the three sub-attributes, after that, assign a value of 10 to the least important sub-attribute, and then assign (equal or) greater values to the second rank and finally to the most important attribute.

Sub-attribute	Value
Travel time	
Waiting time	
Reliability and punctuality of vehicles mode runs come on schedule to the destination	

An example of the Swing (sub-attributes of time)

Considering the goal of the problem (selection of intra-city public transportation mode in the described case), start from the hypothetical worst alternative, in which all sub-attributes are set to their worst possible levels. Then, identify which sub-attribute you would prefer most to change from its worst performance level to its best; you should assign a value of 100 to this sub-attribute. Then repeat this process. Consider the worst hypothetical alternative again and identify the second attribute you prefer to change its level from worst to best and assign a value (equal or) less than 100 to this attribute. You then repeat this for the last attribute.

Alternatives/sub-attributes	Travel time (minutes)	Waiting time (minutes)	Reliability and punctuality of vehicles mode runs come on schedule to the destination
BRT	50	5	High
Metro	45	5	High
Taxi	60	10	Low
Bus	68	5	Medium
Hypothetical worst alternative	68	10	Low
Hypothetical best alternative	45	5	High
Value of sub-attributes			

An example of the BWM (sub-attributes of time)

Considering the goal of the problem (selection of intra-city public transportation mode in the described case), based on your personal preferences, select the most important (best) sub-attribute from the three sub-attributes in the left-hand side cell of the second row (in the designed questionnaire, an option can be chosen from a drop box). After that, express the extent to which you prefer this attribute over the other attributes by using a number from 1 to 9 (in the designed questionnaire, one has access to a complete description of these numbers, and a number can be chosen from a drop-box).

The most important (best) sub-attribute	Travel time	Waiting time	Reliability and punctuality of vehicles mode runs come on schedule to the destination
Travel time, or Waiting time, or Reliability and punctuality of vehicles mode runs come on schedule to the destination			

Considering the goal of the problem (selection of intra-city public transportation mode in the described ca), now select the least important (worst) attribute from the three attributes in the top cell of the second column (in the designed questionnaire, we used drop-box). After that, express the extent to which you prefer an attribute from the first column over the least important attribute by using a number from 1 to 9 (in the designed questionnaire, one has access to a complete description of these numbers, and a number can be chosen from a drop-box).

The least important (worst) sub-attribute	Travel time or waiting time or reliability and punctuality of vehicles mode runs come on schedule to the destination
Travel time Waiting time Reliability and punctuality of vehicles mode runs come on schedule to the destination	